Aalto University
School of Science and Technology
Aalto University
School of Science and Technology
Faculty of Information and Natural Sciences
Degree Programme of Computer Science and Engineering

Klaus Lehtonen

Algorithmic modifications to the emotional content of motion capture data

Master's Thesis

Espoo, August 26, 2010

Supervisor: Professor Tapio Takala

Instructor: Professor Tapio Takala

Aalto University
School of Science and Technology

Aalto University School of Science and Technology Faculty of Information and Natural Sciences Degree programme of Computer Science and Engineering		ABSTRACT OF THE MASTER’S THESIS
Author: Klaus Lehtonen		
Title: Algorithmic modifications to the emotional content of motion capture data		
Number of pages: 75	Date: 26.8.2010	Language: English
Professorship: Interactive digital media		Code: T-111
Supervisor: Professor Tapio Takala		
Instructor: Professor Tapio Takala		

Motion capture has become frequently used in computer games and movies. Since the captured motions can be very realistic, it is also possible to see many emotions in the animations that use the captured material. Having the ability to display a motion with different emotions could be useful for games and storytelling. Capturing one action with different kind of emotions can take a lot of time. To solve this problem modifications that change the emotional content of a motion clip have been developed.

The goal of this thesis was to search and implement ways to modify captured motions in order to change their emotional appearance. Based on the findings of a literature survey, modifications that affect the posture of a character, the length of the motion paths and acceleration of the motion were implemented. A questionnaire with animated motion was made to survey if the modifications really affect the emotions seen in motions.

The results of the thesis show that the changes in posture and in the length of motion paths did help creating the emotions *sad*, *strong*, *excited*, *tired*, *weak* and *afraid*. The emotions *angry* and *relaxed* were not successfully created with the modifications. The modification did also affect the *femininity* and the *masculinity* of the characters in the animations. The results include a list of possible improvements to the modifications that could be a basis for future research.

Keywords: motion capture, human motion, emotion in motion, modification of motion, posture, motion paths

Aalto-yliopisto
Teknillinen korkeakoulu

Aalto-yliopisto Teknillinen korkeakoulu Informaatio- ja luonnontieteiden tiedekunta Tietotekniikan tutkinto-ohjelma	DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Klaus Lehtonen

Työn nimi: Liikesignaalin tunnesisällön algoritmiset muokkaukset

Sivumäärä: 75	Päiväys: 26.8.2010	Julkaisukieli: englanti
Professuuri: Vuorovaikutteinen digitaalinen media		Professuurikoodi: T-111

Työn valvoja: Professori Tapio Takala

Työn ohjaaja: Professori Tapio Takala

Liikkeen tallentamisesta on tullut usein käytetty tekniikka tietokonepeleissä ja elokuvissa. Koska tallennettu liike on hyvin realistista, voidaan sitä käyttävistä animaatiosta nähdä monia tunteita. Mahdollisuus toistaa yksittäinen liike erilaisissa tunnetiloissa voi olla hyödyllinen pelien ja tarinankerronnan kannalta. Yksittäisen liikkeen kuvaaminen monessa tunnetilassa voi viedä paljon aikaa, joten tunteiden lisääminen liikkeisiin muokkaamalla liikettä algoritmisesti on houkutteleva vaihtoehto.

Tämän diplomityön tavoitteena oli löytää, toteuttaa ja arvioida tapoja muokata liikkeitä niiden tunnesisällön näkökulmasta. Tätä varten tehtiin kirjallisuuskatsaus, jonka pohjalta toteutettiin joukko muokkauksia, jotka muuttuvat liikkeen asentoa, liikeratojen pituutta ja liikkeen kiihtyvyyttä. Muokkausten tehokkuutta arvioitiin kyselytutkimuksella, jossa liikkeet esitettiin animaatioina.

Diplomityön tulokset osoittavat, että asennon ja liikeratojen pituuden muokkaaminen sai aikaan tunteita *surullisuus*, *vahvuus*, *innokkuus*, *väsymys*, *heikkous* ja *pelko*. Tunteita *viha* ja *rentous* ei saatu esiin kokeilluilla menetelmillä. Muokkaukset vaikuttivat myös animaatioiden hahmojen *naisellisuuteen* ja *miehisyyteen*. Tuloksena saatiin myös lista mahdollisista parannuksista, joka voi toimia tulevan tutkimuksen lähtökohtana.

Asiasanat: liikkeen tallentaminen, ihmisen liike, tunteet liikkeessä, liikkeiden muokkaaminen, asennot, liikeradat

Acknowledgments

This master's thesis has been made for the Department of Media Technology during a six-month period in the years 2009 and 2010. I thank my supervisor, professor Tapio Takala, for ideas and feedback during the work. I would also like to thank my co-workers Timo, Meeri and Roberto for all the help and the discussions about the thesis. I also greatly appreciate all those people who answered the questionnaire and the actors who helped us to capture the motions for the questionnaire. Lastly I thank Vilja and everyone else who have had to listen to my long lectures about modifications to motion capture data.

Espoo, August 26, 2010

Klaus Lehtonen

Table of Contents

1 Introduction

We observe people around us all the time and see emotions in their actions. We also often show our own emotions in our actions, even when we do not think about showing them. Right from the birth we can already show basic emotions and we learn more all the time. Some gestures that we do are universal to all humans while others are only used by a small group of people. This can cause great misunderstandings if the one showing emotions and the one interpreting them have different ideas about the meaning of a gesture. We do not always want to show our true feelings and we try to deceive other people into believing things that are not true. In some situations it can even be very rude to show your real feelings. (Morris, 1977, pp. 8-34)

If we are looking at a still image of a human, we can often see emotions in facial expressions and the pose of the body. If we would not see the face of a character, but we could see it moving, we might also be able to tell how the character feels. Actors and animators have to think about how emotions can be shown in a way that is clear to the people who see the performances. This can be seen as one of the reasons to study the human motion as it is an important factor in showing emotions.

Studying human motion requires observing humans. Invention of the camera enabled us to record the individual poses in motion with great precision. Creating pictures by hand that are as realistic as pictures from a camera is not a simple thing. This has led animators to use cameras when creating very realistic animated characters. The method is called rotoscoping and it basically means drawing frames of animation on top of a picture taken with a camera. A similar principle can be used in modern motion capture with the difference that many cameras are used simultaneously and the result has three dimensions instead of two. (Menache, 2000, pp.1-2)

Motion capture has been advertised as a replacement for the work of animators, but it has become apparent that this is not true. Motion capture needs someone to act the motions and the actors are limited by laws of physics. This makes capturing motions of a cartoon character or a super hero difficult or in some cases impossible. Also the data that motion capture systems provide is not always ready to be used straight away. Fixing the data can add a considerable amount of costs to using motion capture in animations. (Menache, 2000, pp.37-43)

The use of motion capture has become popular in video games and recently also in movies. This has been possible, because new ways to use and modify motion capture data have been developed. For example fitting the captured data to characters of different sizes has reduced costs as one captured motion could be used for many characters. Similar efforts have been made to change the emotional content of captured motions, but the problem is not solved. Sometimes it can be more effective to have an actor performing one motion in different ways than to make changes in one version of the motion.

The goal of this thesis is to search for ways to modify captured motions in order to change their emotional content. This should be done in ways that are simple to use and have predictable results. The resulting motions must also remain realistic enough to be human motions. To do this a literature survey was made to find out what emotions motion captured data can have and how to modify them. Based on the findings of the survey, a set of modifications was implemented. Post-processing was also necessary to make the modified motions physically plausible. This was only done when clearly necessary. A questionnaire was made to find out if these modifications really affect the emotions seen in the motions. The idea of the questionnaire was to provide data for comparing modified motions to motions captured with skillful actors.

The results show that some emotions can be reliably created with modifications. During the work the importance of taking the restrictions of human body and laws of physics into account became apparent. Motions can easily get unnatural when they are modified. A list of ideas and possible improvements for the modifications was made and this can be a basis for future research.

The following text of this thesis is divided into five parts. The second chapter explores related works which are about how emotions are related to motions and how motions can be modified. The third chapter describes the modifications to motions that were implemented during the thesis. The fourth chapter is about estimating emotional content with a questionnaire. The fifth chapter is discussion about the results. The sixth chapter has the final conclusions of the thesis.

2 Related work

This chapter has several points of view to human motion that were found in the literature. In the first part, we try to understand what gestures are. Then we move on to gestures as part of communication and acting. Next we take a look on the technical side of motion capture and go through a few results on how humans see motions. In order to make a working system we also need knowledge of the mathematics that are involved in the structure of captured motion. After that we take a look on several ways to modify captured motions. The last thing we explore is how to fix small errors in modified motions with post-processing.

2.1 What are human gestures and movements

What distinguishes human movement from the movement of other objects is that human movements often have a meaning, even when they do not physically affect anything in the surroundings. Human movements that carry meaning are often called gestures. In this context, a longer action can be understood as a group of gestures or as a series of movements with an intentional aspect. (Godøy, et al., 2010)

The term 'gesture' has been used to cover many different types of movements. Gestures can be considered from at least three different points of view that are communication, control and metaphor. Communication is involved in gestures during social situations. Human-computer interaction emphasizes gestures that are meant for controlling a device. When music or other cultural topics are at hand, gestures are often understood to be metaphors for something in the related activity. (Godøy, et al., 2010)

Communicational gestures are very common and they can be further divided into five different categories that are iconics, metaphorics, beats, deitics and emblems. Iconics represent something that can be imitated clearly. An iconic gesture can be for example an imitation of a knocking movement while talking about entering a room. Communicational metaphorics are similar to iconics, but they represent more abstract things. A good example would be the raise of hands when referring to “something” in phrase “something happened”. Beats can be hand movements used to emphasize important words when saying something. Deitics are for showing points in space, for example pointing somewhere with a finger. Emblems are gestures with an agreed meaning such as “Hello” or “Ok”. (Godøy, et al., 2010)

When controlling a device, the actual movements are often much more limited than with gestures used in communication between humans. For example when pressing a key on a keyboard everything else than the finger hitting the key is usually ignored by the computer. Recent advances have made it possible to let computers see humans with cameras, but the approach is not without problems. It is quite easy for a human to distinguish hand movements that are meant to say 'Hello' from hand movements used to drive off a fly. For a computer doing the same is not a simple thing. What separates control gestures from communicational gestures is that control often needs physical contact to manipulate something while communication can be empty handed. This separation is not always a very well-defined one. (Godøy, et al., 2010)

2.2 Communication with human gestures

Desmond Morris (1977, pp. 24-34) has studied human behavior and gestures widely. He states that a gesture can be mechanical or meant for communication. Mechanical gestures are for doing physical things like sneezing, but they often also relay a message to people who see them. Gestures that are mainly for communicating can be simply imitations of an object or an action. More abstract gestures can be very symbolic and they can be misunderstood by people from a different culture. A large part of the message that is seen in a gesture comes from the way it is done. For example shaking hands with a smiling person gives a different message than if the person has a very serious face. This happens even if the actual movement of hands is the same. This works also in the other direction. If the person has the same facial expression, but changes the hand shake from *strong* to *weak*, we might interpret the smile as a fake.

The signals we give might be exaggerated or lacking or we might even give controversial signals. When a gesture is exaggerated or lacking it can be interpreted as an attempt to hide the true emotions. When our body language and the actual words we use are in conflict, the body language often wins. For example if you talk of peace but wave your fist at the same time, the violent message might undermine the peaceful one. (Morris, 1977, pp. 112-119)

Face is used for displaying a very large variety of emotions. Parts of the face like the mouth have other important functions besides showing emotions, but for example eyebrows are used mainly for expressing emotions. When a person is not moving at all, the facial expressions are the most important way to communicate emotions. Hands and fingers are also important tools for communication. Hands have universal gestures such as pushing away or inviting someone, but not as many as the face has. Hands and fingers are used often for making gestures that depend on the culture of the actor. (Morris, 1985, pp. 37-108, 145-160)

2.3 Performance theory related to emotional movement

Neff and Fiume (2008) have reviewed a lot of art literature and explored it from a point of view of animation tools. An important principle they found out was that movement in performance is not the same as in everyday life. Performed motions must be very clear so that the audience can notice the message they carry. This can be accomplished by simplification and exaggeration. The less individual movements the audience is shown, the more they will concentrate on the remaining movements. The more a single movement is exaggerated, the easier it is to read. In addition to this Neff and Fiume (2008) divided properties related to motions to three categories that are shape, transition and timing.

Shape means the pose of an actor at a given time. The position of hands, elbows and shoulders can tell how much an emotion is felt. The shape of the torso is important for many poses and plays a big part in dance and sculpture. The posture of the body is often used to show the emotional state and personality of the character. Balance is also important because an actor, that is almost falling down, looks more *excited* than one that stands firmly. The extent and amplitude of actions vary depending on the intimacy or casualness of the actions. For example very private things are done near the body while public actions can happen quite far away from the body. (Neff & Fiume, 2008)

Transition refers to the movement that is done between individual poses. Important things related to transitions are the anticipation and overshoot of actions that emphasize the actual motion. Transitions show if the body of the actor is tense or *relaxed*. Too much tension can be harmful as it can make the actor feel less alive.

Timing movements can be divided into tempo and rhythm. Rhythm is about with the patterns in the motions that are seen especially in dance. Tempo defines the speed which the motions are performed with. An uncertain character might have many different tempo-rhythms, but a determined character might only use one. Another property related to timing is the succession of motion through the body. An example of succession is turning to left and then reaching with a hand to take a cup from a table. If the succession in the example is reversed, the hand would first reach out and then the actor would turn with the hand extended. (Neff & Fiume, 2008)

2.4 Emotions perceived in motions

Lee, et al. (2007) have made effort to map human emotions to measurable physical parameters. They showed movie clips of a device that moved with varying velocity, smoothness and openness to the participants of the study. Smoothness was derived from the acceleration of the device. The openness was derived from the length of the motion path. High openness means that the motion uses as much space as possible. The participants were asked to evaluate the emotion that they perceived in the movie clips. The results of the study show that the velocity and smoothness of the movement correlated clearly with the perceived emotion. Openness of the movement did not have as clear correlation with the emotions as velocity and smoothness. It was speculated that openness might have a clearer effect if it changed during the motion.

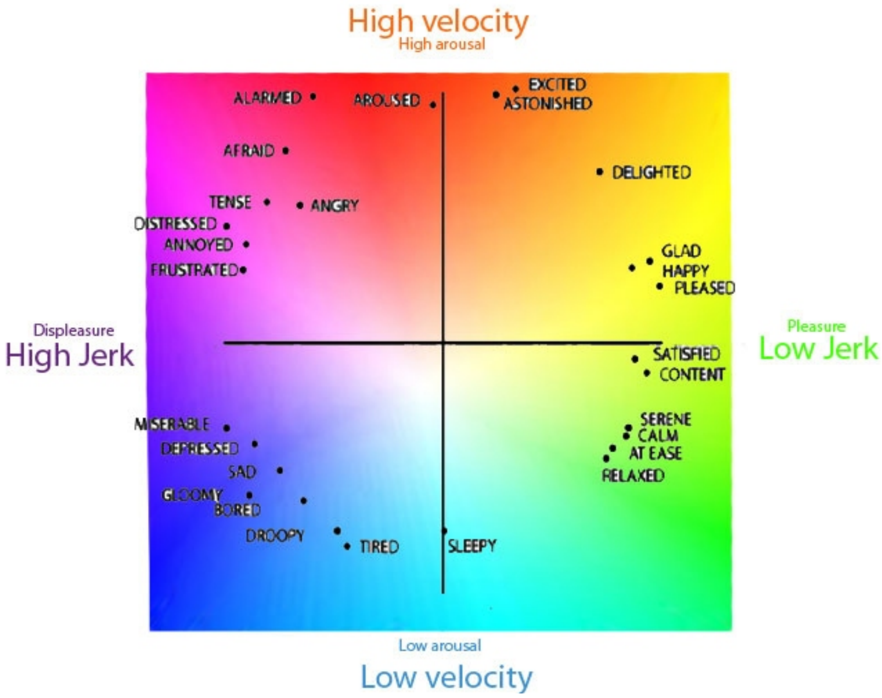


Figure 1: Model used in Mogic for mapping emotions to velocity, jerk and colors. The model was based on previous research. (Neuhaus-Klamer & Dadlani, 2009)

Neuhaus-Klamer & Dadlani (2009) have also made a study related to mapping physical parameters to emotions. They used a device called Mogic to measure the velocity and jerkiness of hand motions that the participants used to represent different emotions. The emotions they included in the study were *anger*, *disgust*, *enjoyment*, *fear*, *sadness* and *surprise*. Based on the previous studies in the field, they assumed that the emotions could be mapped to physical movement as in figure 1. The results based on data gathered with Mogic show that the assumed model was not accurate (figure 2). Especially *enjoyment* and *anger* were much harder to distinguish from the data than the original model suggested. Video recordings of Mogic in use showed that when expressing positive emotions the participants pointed Mogic up and when expressing negative emotions they pointed it down. So while *enjoyment* and *anger* were not clearly separated by velocity and jerkiness, they can be distinguished if the direction of Mogic is also taken into account.

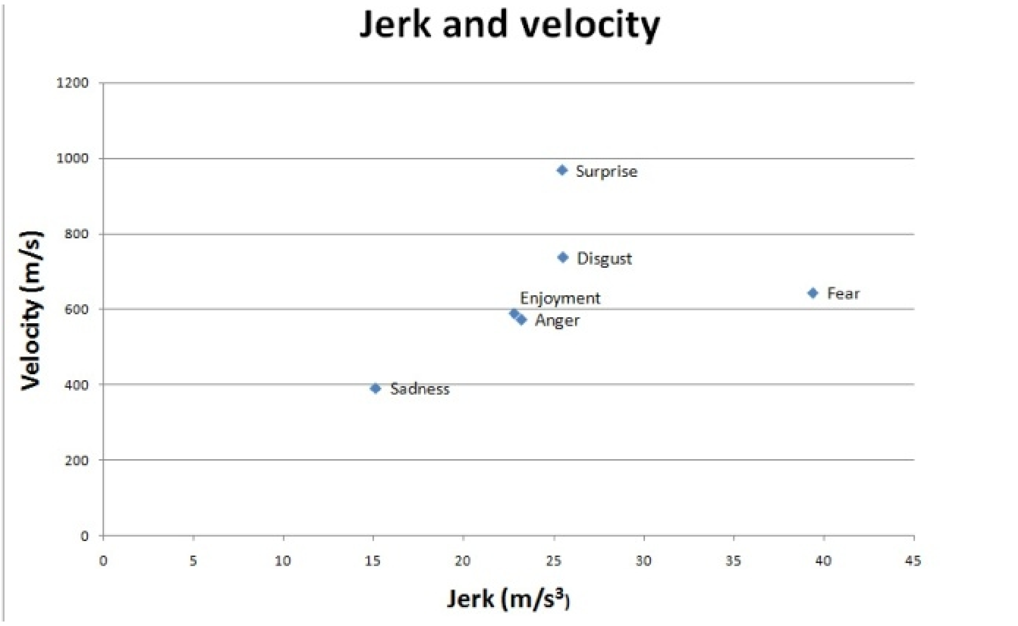


Figure 2: Mapping of emotions to velocity and jerk based on the data gathered with Mogic. (Neuhaus-Klamer & Dadlani, 2009)

2.5 Motion capture in performance animation

Performance animation refers to the whole process of bringing life to an animated character. Motion capture can be used as part of performance animation. To make a complete performance animation one needs to model the character and map the motion to the character. Motion capture is a good way to produce realistic looking performance, but it is limited to what the actors are able to perform. (Menache, 2000, pp.1-2)

Motion capture can be as simple as drawing a character based on a performance recorded on film. This technique is called rotoscoping. More advanced methods for collecting human motion can be classified into three types of systems. An outside-in system collects data from the human motion with sensors that are not attached to the body. This can be for example a camera system that tracks markers placed on the actor. Inside-out systems have sensors placed on the body. The sensors can then measure a signal that is generated by devices outside of the body. This principle is used in many electromagnetic systems. Inside-in systems do not rely on any external parts, but measure the motions directly from the joints of the actor. This method is used in electromechanical suits. (Menache, 2000, pp.2-3,14-31)

Entertainment is a major use of motion capture, but many motion capture systems have been developed for other purposes such as medicine, sports and even as evidence in legal cases. In entertainment, the main applications of motion capture are video games, television and feature films. All these set different requirements for the accuracy, cost and realism of the resulting performance. For example a video game might not require as realistic motion as a feature film. Similarly, an error of one centimeter might not matter in a feature film, but could be a matter of life and death in a medical operation. (Menache, 2000, pp.32-36)

2.6 Motion signal from a structural point of view

When animating a human we need to decide what the basic structure of the motion data is. If only large motions are needed and the small movements of muscles and skin are ignored, a skeleton based approach can be a good solution. A skeleton consists of bones and joints. The bones are assumed to have a constant length and the joints are assumed to be purely a rotation. The bones could be represented with start and end points, but that way when editing the motion we would need to handle a lot of constraints to stop the bones and joints from stretching and breaking. A better way to represent the skeleton is to define the bones relative to their parent bones and use the rotation of the bones instead of the start and end points. This way we have a transformation hierarchy that implicitly takes into account a large part of the constraints in the skeleton. A transformation hierarchy is also beneficial because many 3D graphic APIs such as OpenGL support automatic traversal of the hierarchy. (Grassia, 2000)

If a hierarchical skeleton structure has been selected, we need to decide what the best way to represent joint angles is. To do this we need to know what we are going to do with the motion data. If we only want to show the motion, almost any representation is good enough. Quite often we need to change the motion with, for example, signal processing algorithms or by blending the motions with other motions. Any motion editing sets conditions for the representation of the joint rotations. (Grassia, 2000)

We can use Euler angles for the joints, but this can have many unwanted side effects. Using Euler angles means defining rotations in a predefined order around the three coordinate axes. This can make the motion to go near a singularity and cause signal processing algorithms to work in unpredictable ways (Grassia, 2000). Interpolating between two Euler angle rotations does not always result in natural looking motion. Euler angles can also end up in a situation called the gimbal lock where we lose one degree of freedom as a result of two co-aligned axes (Shoemake, 1985).

Another way to represent the joint rotations are unit quaternions. A quaternion is a set of four scalar values that specify a rotation of a specified angle around an axis. Two quaternions cannot be interpolated by directly, but we must use for example spherical Bezier curves (Shoemake, 1985). A good side of quaternions is that they are safe from gimbal lock that is a problem with Euler angles (Shoemake, 1985). A downside is that multiple quaternions cannot be combined with standard statistical tools, but require more advanced mathematical methods such as radial basis functions (Grassia, 2000).

When we have selected a way to represent the skeleton for one posture we still need to decide how to handle motions as functions of time. Motion capture systems give the motions as a discrete set of values evenly spaced in time. Animation created by hand consists of key frames and a well-defined way to interpolate the in-between frames. The key frames can be much more widely spaced than the frames in motion recorded with motion capture systems. The key frames may not be evenly spaced. Difference in the amount of data between the two representations can be very large. That means that part of high frequency components are lost if one wants to fit motion capture data to a set of key frames (Grassia, 2000).

2.7 Modifying motion capture data

In the following parts, many existing methods for modifying motion capture data are presented. While the methods themselves are explored from many points of view in the publications, their effects on the emotional content have often been only narrowly tested. Mathematical details of the methods are only included if they are relevant to the motion modification system implemented in this thesis.

2.7.1 Motion warping

Motion warping requires an animator to define new keyframe constraints inside the motion. The flow of time and joint angles around the new key frames will then be changed to fit the constraints. This technique is similar to the conventional keyframing and therefore it can easily be used by animators. Motion warping preserves fine details of the motion in high frequencies. This helps to preserve the aliveness of the motion if too extreme changes are not made. (Witkin & Popovic, 1995)

Joint angles are warped independently which means that we only need to consider a single joint. Animator supplies the warping system a set of constraints that define what the joint angle should be at a chosen time. The animator can also supply time constraints that define a time t_1 from original motion that is mapped to time t_2 in the new motion. Time warping can be done by constructing a spline that smoothly interpolates the time warp constraints and by fetching the corresponding angle from the original motion according to the interpolated time. (Witkin & Popovic, 1995)

Motion warping is not as low level as editing each key frame individually, but it still does require a lot of human expertise. A problem like raising the arm of a tennis player to make the racket hit the ball can be fixed with the technique. The technique does not help in larger modifications like changing an energetic walk calmer.

2.7.2 Interactive editing of motion style using drives and correlations

Neff and Kim (2009) described a system for editing captured motions and key framed motions by using motion drives and correlations derived from the motion. A motion drive is a linear correlation between two spatial positions in the skeleton. A motion drive can be for example the distance of a wrist joint from the torso. The idea behind motion drives is that they describe shaping over time instead of a single frame. In addition to spatial positions, correlations can be based on joint rotations or a mix of positions and rotations.

The system calculates the most frequently used correlations automatically when a motion clip is loaded. The animator can also define new correlations. When editing the data the animator changes the scales and shifts of the drives and correlations. The system then produces the changes with inverse kinematics. Various changes in the posture and style of a movement such as turning a regular walk to a limbing walk has been done with the system. It can also be used to fix sliding feet that often appear in the changed motions. (Neff & Kim, 2009)

The system Neff and Kim (2009) have described is more suitable for changing the emotional content of motions than the motion warping described earlier. Editing motion drives and correlations requires a lot of work and talent from the animator and cannot be used without human supervision. This makes the system unsuitable for situations where the emotional status of the character can change freely and is not one from a predefined set.

2.7.3 Style components

Shapiro et al. (2006) used independent component analysis to separate two joined motion clips to style components that are also motion clips. This allows the animator to take components from one motion and put them in another motion to produce a new motion with a new style. The method is completely visual for the animator and requires no knowledge of key framing, frequency bands or statistical analysis.

The idea is to find the things that make the two versions of a motion different. To make the method work the two versions must be combined. This simply means that one motion is copied to the end of the other motion. Then the combined motion is decomposed to components with independent component analysis. After the decomposition the two versions are treated separately. Next the animator selects the interesting components and the system can then join the selected components into a new motion. For the joining to work the two versions must be syncronized and made the same length with time warping. The new motion might also need post processing because the method can produce high frequency vibrations and make the feet slide. (Shapiro, et al., 2006)

The technique is interesting as it can be used without much animation experience. However, to use the technique a large collection of captured motions would be required. The technique is not fully reliable, therefore using it in real-time applications to produce different motions is not realistic. More thorough evaluation of the possible types of style changes with the technique could also be useful as the definition of a style is not very clear.

2.7.4 Emotion from motion

Amaya et al. (1996) presented a way to make emotional transforms to motions. Their approach was to capture one motion performed with different emotional states such as *sad*, *neutral* and *angry*. Then they proceeded to calculate a transform that holds the differences between the emotional and the *neutral* motions. The transform could then be applied to a new *neutral* motion to make the motion more emotional. The transforms are intended to capture differences in speed and spatial amplitude. Speed is closely related to timing and spatial amplitude is related to the range of the motions.

In order to calculate the speed transform it is necessary to divide both the emotional and the *neutral* motion to basic periods. The periods are bounded by frames where the velocity of the end of the limb is changing direction or the joints of the limb are between extension and flexion. For example in a drinking motion the periods could be "hand to the cup, cup to mouth, cup down, hand back". All the basic periods are processed separately. The speed transform for a single period can then be done by modifying the distribution of the frames to make it similar to the distribution in the emotional data as shown in figure 3. (Amaya, et al., 1996)

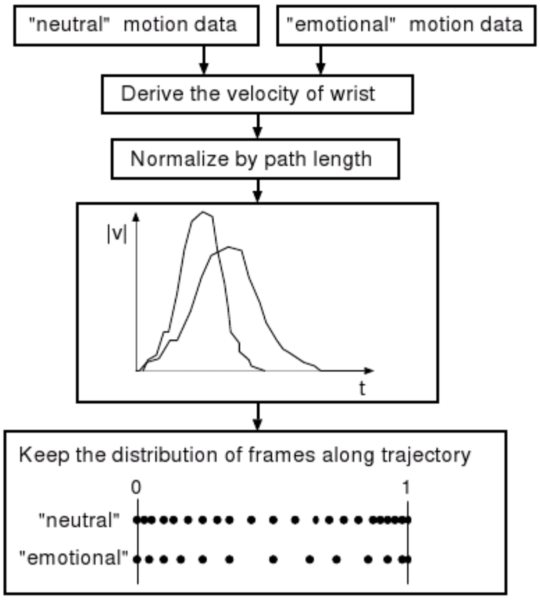


Figure 3: Algorithm to obtain a speed transform of a basic period in motion capture data (Amaya, et al., 1996).

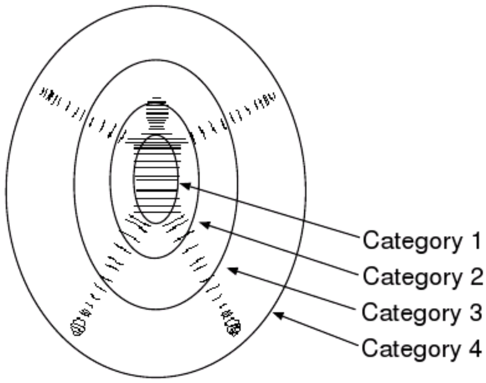


Figure 4: Division of the human body for the spatial amplitude transform (Amaya, et al., 1996).

For the spatial transform the human body is divided into four categories based on the distance from the center of the body. The categories are shown in figure 4. This is needed because range of the motion in the categories differs greatly. The actual transformation is done separately to each basic period of each joint.

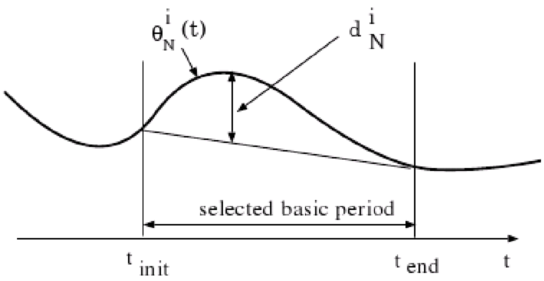


Figure 5: Intensity factor of spatial amplitude d^i can be measured from a motion signal of a joint (Amaya, et al., 1996).

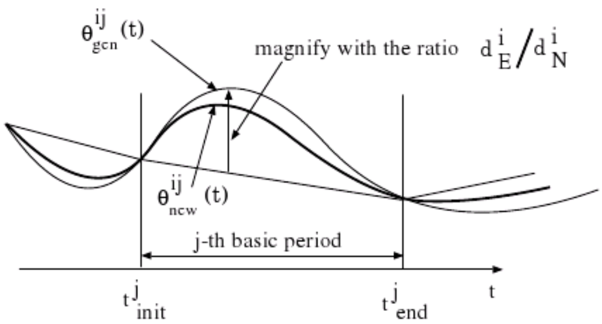


Figure 6: Transformation of spatial amplitude creates a new motion signal for the joint (Amaya, et al., 1996).

The idea of spatial intensity d^i is to represent the maximal displacement of the joint during the basic period as is shown in figure 5. When the spatial intensity of the emotional motion d^i_E and *neutral* motion d^i_N is known, it is possible to transform the *neutral* motion as it is shown in the figure 6. (Amaya, et al., 1996)

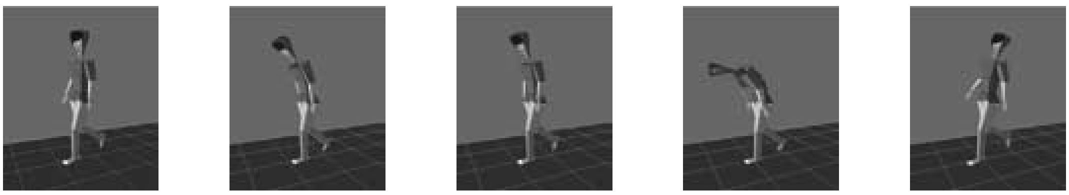
At first Amaya et al. (1996) considered the high frequency content of the motions to be important. Still, after the analysis of motions in the frequency domain they noticed that neither emotional nor *neutral* motions had any significant components in frequencies higher than 10 Hz. They also noted that in an animation that has a frame rate around 30 frames per second a lot of the high frequencies are lost. Therefore high frequencies cannot be very important to emotions that are visible in an animation.

The method which Amaya et al. (1996) used to modify motions does not take into account phase shifts between the joints. However, they reported finding differences between the phase shifts in *neutral* and emotional motions. This made them consider modification to phase shifts as a possible direction for future research.

As a whole the system described by Amaya et al. (1996) has a lot of potential as it clearly defines the modifications that affect the emotional content of a motion. It might be possible to adjust the system to work in a real-time environment like games. However, the requirement of a captured both *neutral* and emotional motions can reduce the value gained from the system.

2.7.5 Fourier principles for emotion-based human figure animation

Unuma et al. (1995) used the Fourier expansions of motion capture data to interpolate and extrapolate cyclic human motion. This enabled them to get styles such as *tired* and brisk from captured motions and to create new motions with the styles as is shown in figure 7. The method works only for cyclic motions like walking or running and not for non-periodic motions.



a. *b.* *c.* *d.* *e.*

Figure 7: Motion captured motions walk (a) and tired walk (b) were used to generate slightly tired walk (c), very tired walk (d) and walk with negative tiredness (e) (Unuma, et al., 1995).

The system shows that modifying cyclic motions can be done quite effectively, but cyclic motions are only a small set of all human motions. Also, because the system only considers cyclic motions, non-repeating details in motion might be lost. An important thing is that it can be used to create motions in real-time.

2.7.6 Style translation for human motion

Hsu et al. (2005) have worked on a way to change the style of a motion with a simple linear operation to each frame based on the neighboring frames. Their approach needs two motions with different styles paired together as input and desired output. Then the system can be trained using the two motions and an iterative optimization method. The result of the training is a linear time invariant model that represents the change in style between the motions. The model can be used to change the style of a motion simply by filtering the motion with the model. This makes the system good for situations that need fast processing like games and live performances.

When testing the system Hsu et al. (2005) noticed that normal walking can be reliable transformed to other styles. However, if the walk had parts that were different from normal walking like stepping over an obstacle, the output of the system was unpredictable. The irregularities vanished after the deviating parts of the motion ended. To solve this, heuristic detection of deviating parts was used and the original motion was displayed instead of the one with transformed style in the detected parts.

2.7.7 Motion signal processing

Bruderlin and Williams (1995) examined many methods of modifying motion capture data as a signal. They tried out methods on motion signals that were based on joint angles or joint coordinates. The methods were considered complementary to earlier key frame editing and procedural animation techniques.

Multiresolution filtering is a way to divide motion signal to frequency bands which can be amplified as desired and then merged back to one motion. The number of bands (fb) that can be generated is relative to the number of frames (m) in the signal:

$$if\ 2^n \leq m \leq 2^{(n+1)}, then\ fb = n \tag{1}$$

The frequency bands can be generated by successive low pass and band pass filtering, but Bruderlin and Williams (1995) decided to go with a more simple solution. They kept the bands at the same length and expanded the filter kernel ($w_{X)}$ by adding zeros between its values as follows:

$$w_1 = [c\ b\ a\ b\ c];$$

$$w_2 = [c\ 0\ b\ 0\ a\ 0\ b\ 0\ c]; \qquad (2)$$

$$w_3 = [c\ 0\ 0\ 0\ b\ 0\ 0\ 0\ a\ 0\ 0\ 0\ b\ 0\ 0\ 0\ c];\ \text{etc.},$$

The multiresolution algorithm is relatively simple and works simultaneously for each channel in the signal. First you must get the low pass filtered bands G_0 to G_{fb} where G_0 is the original signal. Then the pass band versions L_X can be generated by subtracting G_{k+1} from G_k. After you have the pass bands L_X you can multiply them by the desired gain values. To reconstruct the signal you only need to sum the G_{fb} and all pass band signals L_X.

Bruderlin and Williams (1995) found out that adjusting the gains of motion defined with joint coordinates resulted in a cartoonlike squashing and stretching. When using joint angles they noted that adjusting middle frequencies resulted in exaggerated motions and increasing high frequencies added the nervous twitch to the motions. They also noted that adjusting the gains could easily break joint constraints and make feet slide or hover above ground.

To fix the joint constraints such as preventing the knees from bending forward Bruderlin and Williams (1995) proposed motion waveshaping. This can be done by setting hard limits to joint angles such as if angle x is larger than y, limit x to y. Hard limits can make motions stop suddenly when the limit is reached. Soft limits do not have that problem. Soft limits can be made by mapping values that are over the limit to values that gradually approach the limit.

When blending two motions Bruderlin and Williams (1995) found out that the motions need to be on the same phase for the blend to work. For example when blending two different walks without any effort to make them similarly paced the result might look like a mix of uncoordinated motions. To fix this, time warping is needed to make the motions similarly paced.

Motion displacement mapping is a technique that Bruderlin and Williams (1995) presented that allows animators to easily fix errors and make local changes in the motions. The method relies on fitting splines to the motion capture signal. Then the splines can be edited to make changes in the motion. This method is similar to reducing the amount of frames in the signal and editing those in the same way as in key frame editing.

2.7.8 Comparison of the modifications in literature

The techniques that were presented in previous chapters have very different approaches to modifying human motion. Some of the techniques require a lot of human expertise to work while others rely on a system trained with motion clips to handle all the small details. Another way to classify the techniques is to divide them into those that see human motion as a statistical process and to those that try to modify the physical properties of the motions.

From a technical point of view the techniques handle motion data very differently. Bruderlin and Williams (1995) consider all the joints to be quite independent, while Neff and Kim (2009) see the poses that the joints form together as the most important thing. Shapiro et al. (2006) also consider the joints together, but then divide the motion into independent components that describe the motion in more natural parts than the movement of the joints. Bruderlin and Williams (1995) and Unuma et al. (1995) both consider that the frequency representation of the motion is very important. Unuma et al. (1995) take this so far that they only consider the frequency content limiting themselves to the cyclic motions.

It is clear that the earlier research is not a very tightly knit field. When building a system that modifies motions, you simply cannot rely on joining the earlier methods. The large number of different methods demonstrates the difficulties that are related to human motions and indicates a niche for new research.

2.8 Fixing errors in modified motion

Raw motion capture data can have many irregularities caused by badly calibrated sensors or markers being occluded by the body of the actor or another object. Often the data can be salvaged, but the repairs might add more small errors to the data. Usually, the raw sensor data is not used, but a skeleton of the actor is estimated from the data. The estimation step might also produce small errors to the data. These errors might sometimes go unnoticed, but for example errors in foot plants are easily seen by humans. Foot plants are moments when a foot is in a fixed position on the ground. A fixed position is also required if a human leans on an immovable object with hands. Keeping the positions fixed requires high frequency details in the motion signal to look realistic. These details might be lost especially when signal processing techniques are used. (Gleicher, et al., 2002)

If we want to fix errors in a motion, we need to define constraints that tell when the motion is not acceptable. Gleicher (1998) lists seven ways to get suitable constraints:

1. A parameter value is in a certain range. This is useful for joint rotations.
2. An end of a limb is in a required position. This is useful for foot plants.
3. A point in the character is in a certain region. For example a foot is above the floor or a hand is not inside a wall.
4. A point in the character remains in the same position. This is useful for preventing skidding.
5. A point of the character is following a path of another point.
6. Two points are at specified distance. This can be used for example when a character is carrying a fixed-sized object.
7. A vector between two points has a certain orientation.

These constraints can be easily calculated from a readily available motion data. More difficult problem is how to modify a motion if it breaks a constraint. If the goal is to fix foot plants or fit a motion to a character of a different size, inverse kinematics can be used to find suitable joint angles. If a good solution cannot be found by adjusting joint angles, the length of the bones might also need temporary adjustments. (Gleicher, 1998; Gleicher, et al., 2002)

A physics-based approach to the correction of motion data is also possible. The first step in a physical model is the mass distribution of the character. This is not usually part of motion capture data, so a model for an average human is used. The mass distribution can be estimated for all body parts with simple elements like boxes and cylinders. With the mass distribution and the recorded data it is possible to calculate the forces and torques in the motion. If a force or torque is too high in a part of a motion, we can quite reliable say that the part is unrealistic and needs fixing. The physical model can also be helpful when fixing foot plants of a character. With a physical model it is possible to calculate the friction between the foot and the floor and use this information to make accurate fixes. (Xiao, et al., 2005)

3 A system for modifying emotional content in motion capture data

This chapter has a technical point of view. First the goals for the system are described. Then the basic structure of the motion signal and the skeleton used for the character are presented. Next the modifications and the necessary post-processing are explained. Last parts of the text are about using the modifications together.

3.1 Goals for the system

Based on the literature survey we defined the requirements of the system to modify emotions in motion capture data. The system should be predictable, easy to use and simple enough to work in an interactive environment. Also the system must produce physically plausible motion or motion that can be easily fixed to look realistic.

The goal is to modify motions and not to create new ones from scratch, therefore we decided to concentrate on modifying the style of the movements while preserving the action seen in them. This leaves out modifications to emotions based on symbolic gestures because they would add new actions. Another thing to limit the possible range of emotions is that capturing the fingers and the face was not supported by the motion capture system that was available.

If the modification system is predictable, it must be possible to guess the results based on the input of the system. One input must be the motion that we want to modify. Other inputs define the changes in the style of the motion. In the literature the style is sometimes defined by motion clips. This is not a very clear way to define the style because everyone does not see all motions in the same manner. Another way is to give the system a small set of numbers that represent some properties in the motion. This approach can be clearer than using a motion clip to define a style, but selecting the right properties might be difficult.

We decided to define the style of a motion as the product of the posture of the character, the length of the motion paths and the acceleration of the motion. These parameters do not have direct correlations with any emotions and some emotions might not be possible to produce by editing them. However, these three parameters can be clearly defined and they are considered important in performance theory. After consideration it was decided that the best way to proceed was to build a system to modify the three parameters and then test the effectiveness of the system with people viewing the motions.

3.2 Tools

The system used for recording the motions was Optitrack Full Body Motion Capture system that is a commercial product of Natural Point Incorporated. The system consisted of twelve FLEX:V100R2 cameras. The cameras have infrared LEDs (light-emitting diodes) around the optics as seen in figure 8. The infrared light is reflected by the markers and so the cameras see the positions of the markers in the scene. The markers are not different from each other as can be seen from figure 9. Detecting the position and rotation of a rigid object requires at least three markers. The cameras were used with the Arena software version 1.5 that was running on a Windows XP laptop. The setup enabled us to get motion performed by an actor in Biovision Header (*bvh*) format. The skeleton of the actor in one frame is represented with coordinates and the rotation of the hips and 18 rotations of the joints. The rotations in *bvh* file are given as angles around Y-, X and Z-axes. Bone lengths between the joints are included in the header of the *bvh* file. The cameras are capable of capturing 100 frames per second that forms hundred lines of numerical data to the file.

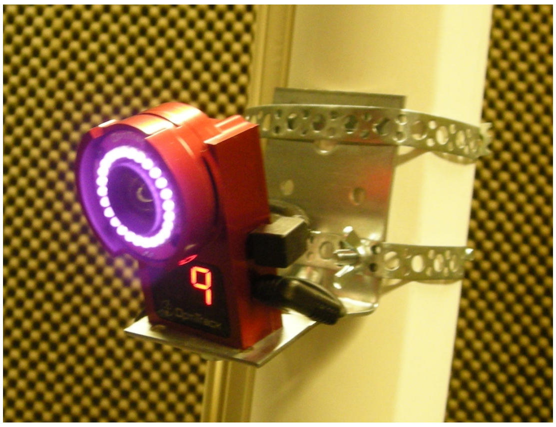


Figure 8:The FLEX:V100R2 camera that was used to capture the position of the reflective markers.

Figure 9: The suit used in the capture of the motions. The reflective markers are lighted by the flash of the camera.

All of the coding and motion editing was done on a Linux system. Tools used for modifying and viewing the motion capture data were Matlab 7.9.0, GNU C++ compiler, standard C++ libraries and OpenSceneGraph library version 2.8.2. Programs coded with C++ and the OpenSceneGraph library were used for viewing the motion data in three dimensions, changing the format of the motion data and fixing sliding steps in some motions. Matlab was used for visualizing the motion signals with plots and making all the modifications of the motions that required signal processing.

3.3 Structure of the motion data

In the literature survey it was found that most popular way to represent the human skeleton was with fixed length bones and varying joint rotations. The joint rotations were usually defined with Euler angles or quaternions. A less popular way was to represent the skeleton with the coordinates of the start and the end points of the bones.

Modifications to the data with signal processing techniques set two requirements for the skeleton. First requirement is that making changes in the data must not squeeze or stretch the character. In other words the bone lengths must not change. Second requirement is that a small change in the pose of the skeleton must result in a small change in the numbers that represent the pose.

A few tests were made with skeleton representation that was based on the start and end coordinates of the bones. Using coordinates allows simple mathematics to be used. A downside of a coordinate based representation is that with only start and end points of the bones some rotations are lost. The start and end points define the direction of the bone, but not the rotation around the bone. When viewing the motions this representation did work well, but when making any modifications to the data the bone lengths break very easily. It was clear that realism would be lost if coordinates were used in the modifications. The representation might be useful if cartoon like effects are desired, but that was not the goal.

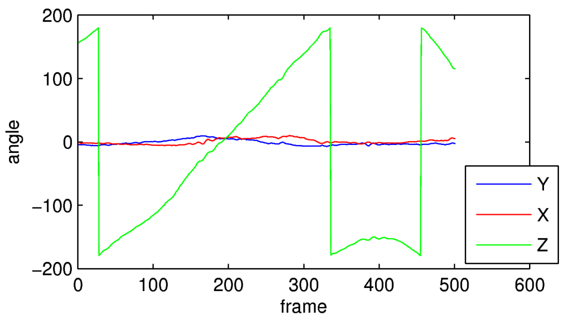


Figure 10: The euler angle representation of a smoothly rotation rigid object. Notice the jumps of 360 degrees.

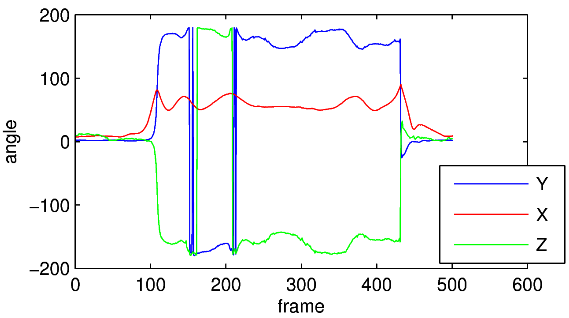


Figure 11: Gimbal lock found in a knee joint when bending the knee 90 degrees. Y and Z-axes have become locked in mirrored angles and swap around near frame 200.

The *bvh* format used fixed bone lengths and joint rotations with Euler angles. Modifications to the data were tried out with this skeleton representation. The format obviously did not have problems with bone lengths, but small changes in bone positions did not always show as a small change in the joint angles. The joint angles sometimes jumped from +180 degrees to -180 degrees as seen in figure 10. Jumps of 360 degrees can be fixed, but a more serious problem comes if the angles go near a gimbal lock situation. A gimbal lock in a knee joint is shown in figure 11. A gimbal lock makes the joint angles behave very unpredictably and that can easily break many signal processing techniques.

Quaternions are not prone to the gimbal lock, therefore they were considered a possible alternative to angle based joint representation. Figure 13 shows how a quaternion behaves when the angle representation would be in a gimbal lock. Quaternions are supported very well by the OpenSceneGraph library. This led them to be used a lot when programming tasks related to viewing motion data. Quaternions were not used when making modifications to the motion data, because quaternions cannot be modified with simple Euclidean geometry and because they are ambiguous. The ambiguous behavior is caused by the fact that a quaternion and its inverse represent the same rotation. This can cause similar problems as the 360 degree jumps in angle representation in figure 12. If the jumps had been fixed, the quaternions would probably have been usable in the modifications done by signal processing.

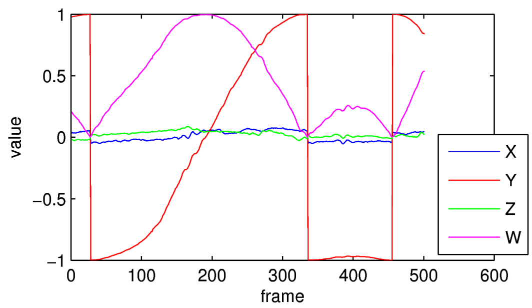


Figure 12: The same rotation as in figure 10 in quaternion format. Notice that in addition to the large jump in Y-component, there are smaller jumps and changes of direction in all other components.

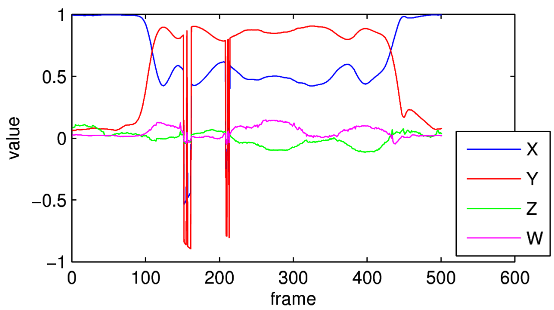


Figure 13: Gimbal lock of figure 11 converted to quaternion format. Notice the absence of locked and swapped components compared to figure 11.

Motion signal representations that were found in the literature survey were not as robust and simple as we had hoped, therefore we had to search for new alternatives. A suitable solution was found by using a coordinate based representation of joint angles that was separated from bone lengths. By using two vectors that have three dimensions and a 90-degree angle between them, it is possible to define the direction of a bone and the rotation around that direction. Length of the vectors can be ignored as they are only used for directional information. The direction vector representation of joint angles does not suffer from jumps as is shown in figure 14, because the vectors do not have any ambiguity in them. The representations can also handle gimbal lock situations as is shown in figure 15, because gimbal lock is a problem only when using angles.

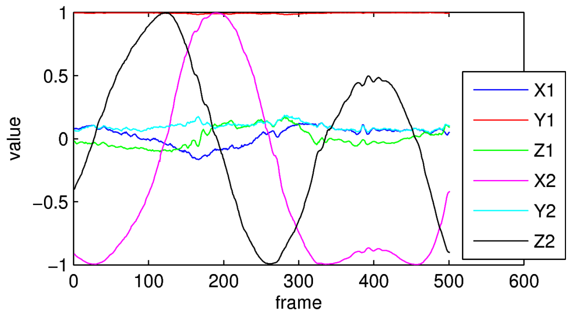


Figure 14: The same rotation as in figure 10 in direction format. Compared to figures 10 and 12 the curves are smoother.

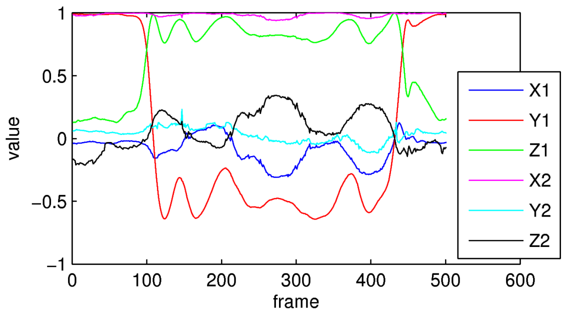


Figure 15: Gimbal lock of figure 11 in direction format. Notice the absence of jumping compared to the figures 11 and 13.

Compared to quaternions the direction vectors allow much simpler mathematics to be used, therefore the direction vectors were a good choice from signal processing point of view. There is one very important difference between the direction format and the angle based format. When multiplying an angle with a big enough number the rotation goes around a 360-degree cycle. Because all the joints have constraints to their rotation, the multiplication will eventually break the constraints. However, when multiplying a direction vector the length of the vector changes, but the direction remains the same. This makes direction vectors less prone to breaking joint constraints.

3.4 Methods for modifying motion capture data

The following sections describe the methods that turned out to be useful when modifying motion capture data. Modifying the length of motion paths, posture and the speed of the motion were thoroughly examined and implemented. Other techniques like adding noise to the motion signal, adjusting the overall flow of time in the animation and modifications based on the independent component analysis were also tried out. These techniques were left with less attention because of difficulties in implementing them or because they did not seem to affect the emotional content of the motions when viewing the modified motions.

3.4.1 Length of motion paths

To change the length of the motion paths, multiresolution filtering described by Bruderlin and Williams (1995) was used. The technique was implemented with Matlab exactly the way it is described earlier in the literature survey. With multiresolution filtering we can take a motion like in figure 17 and exaggerate its motion paths to look like in figure 18. We can also use it to diminish a motion. The technique handles all the components of the motion signal separately and it can be used with almost any motion signal format. The formats that were thoroughly tried out had a skeleton with constant bone length and the joint rotations were represented with angles or direction vectors. The best results were produced with direction vectors, an example of the resulting signal is shown in figure 16. A few tests were also done with joints represented with quaternions. The quaternions seemed to work quite similarly to the direction vectors, but the tests were not very thorough as the direction vector format gave good enough results.

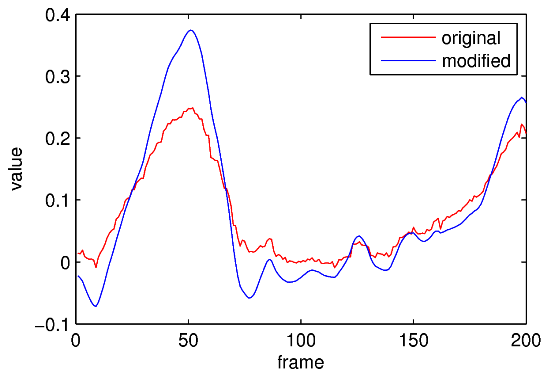


Figure 16: Original and exaggerated motion produced with multiresolution filtering. The signal is one of the six components that are used to represent the rotation of the joint signal from a left elbow in the direction vector format.

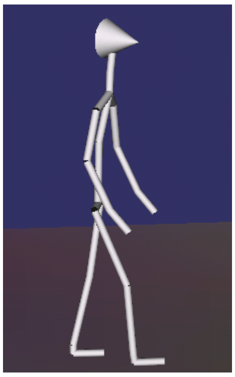


Figure 17: A frame from animation with original motion capture data.

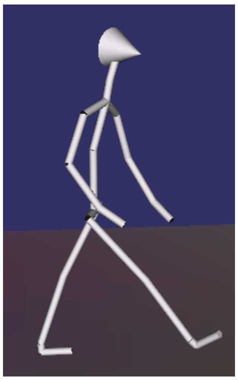


Figure 18: The same frame of animation that is seen in figure 17 with longer motion paths.

The idea of multiresolution filtering is that the original signal is divided into frequency bands that can be processed individually. Before dividing the signal to the bands, the signal was lengthened from both ends with the first and last values to ensure that no sudden jumps appear in the signal. After division into the bands is ready, the lengthened parts can be cut out. The algorithm of the modification as pseudocode is the following:

Calculate the amount of bands per channel *fb* allowed by the filtering;
For each channel of each joint in the motion do
    Double the length of the signal by adding repetitively the first value to the start and the last value to the end of the signal;
    Create low pass filtered versions *G(1)* to *G(fb)* by consecutive low pass filterings;
    Create pass band versions *L(x)* by taking *G(x)* minus *G(x+1)* for each band;
    Multiply the values of each *L(x)* by the desired gains;
    Sum all the multiplied pass bands sample by sample to get the new signal;
    Removing the lengthened parts from the start and the end of the signal;
Done;
	(3)

Figure 19 shows an example of what the frequency bands look like. Any signal processing operation could be done to the frequency bands. However, we only need to do a simple multiplication to the bands to make the motion paths exaggerated or reduced. The bands can be roughly divided into low (under 1.0Hz), middle (between 1.0 and 12.5 Hz) and high frequencies (over 12.5 Hz). The lowest frequency band, that is the red one in figure 19, can have an offset from the zero level. The other bands have almost a zero average in the long run.

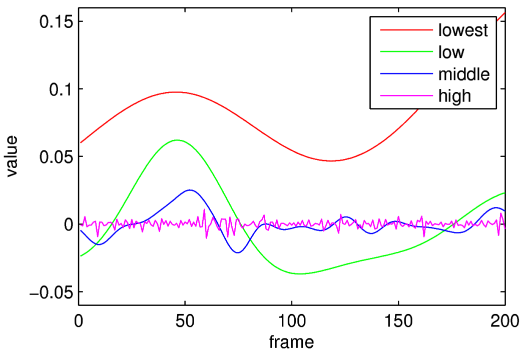


Figure 19: An example of the frequency bands derived from the original signal shown in figure 16.

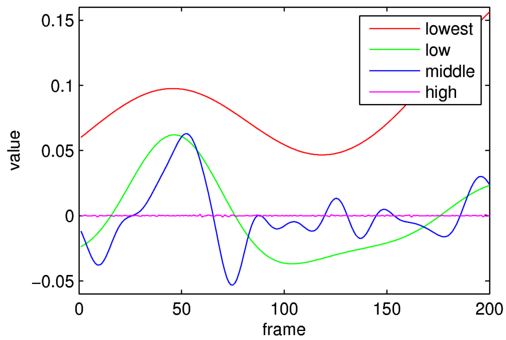


Figure 20: Same frequency bands as in figure 19 multiplied individually with constants. We get the modified signal shown in figure 16 by summing all the multiplied bands. The high frequency band (magenta) is reduced and the middle frequency band (blue) is amplified.

Tests were made with a few recorded motions to find out how multiplying the bands affect the visible motion. When amplifying the high frequencies all the noise and small errors in the signal seemed to become more visible. When the high frequencies over 12.5 Hz were heavily reduced the motion looked almost identical to the original. These observations lead to the decision that high frequencies were always toned down in the final version of the modification.

The middle frequencies between 1.0 and 12.5 Hz are the most important when adjusting the motion paths. Amplified middle frequencies make the motion paths look longer. The effect of amplifying a middle frequency can be seen when comparing the blue signal in the figures 19 and 20. Too large amplification can make the motion unrealistic especially if angle representation is used for the joint rotations. When using direction vectors for the joints, the joints were less prone to bend too much. When making the middle frequencies reduced the motion paths became smaller. Reducing the middle frequencies worked equally well in both joint formats.

Changing the low frequencies under 1.0 Hz caused very different effects when done with angles compared to direction vectors. With the joint angles the lowest frequency band must be kept unchanged, because it changes the average rotation of the joints. When using the direction vectors the low frequencies affect the length of the vectors. This does not cause any visible problems as only the direction of the vector is used. The ratio between the multipliers of the low frequencies and the higher ones is the thing that matters. A low multiplier in low frequencies has the same effect as keeping the low frequencies unchanged and having a high multiplier in the middle and high frequencies.

Besides the signals for rotations, the skeleton also has signals for the coordinates of hips. At first they were also modified in the same way as rotations, but the resulting motion did not always look very good. Better results were produced when the coordinates were multiplied with the same number as the joint middle frequencies without any division into bands. This way the center of the character moves about in the way the legs make it move. However, this approximation is not enough, because humans can very easily see if the feet of the character are moving when they should be locked to the ground. To fix this problem post-processing of the motion was needed. In the end, the coordinates of the hips were produced with post-processing and the modification was not used on them.

The division of the signal to frequency bands requires quite a lot of processing. For this reason alternate ways to get the same effect were also explored. It is quite simple to produce a digital filter that has a similar effect on the frequencies as the multiresolution filtering. When experimenting with the alternatives it became quickly obvious that most of the filters are not usable as they add phase shift between the frequencies. Phase shift can make the resulting motion look very unnatural as the slow and fast motions are not timed correctly. While it is possible to make an alternative filter that adds no phase shift, it is not easy to make the filter more efficient than multiresolution filtering. In the end the ideas of finding alternative filters were abandoned.

The lack of phase shift makes interpolating between different motions possible. If we have one exaggerated or reduced motion, we can get all the motions between it and the original motion with a simple linear interpolation. This can be a very useful property as linear interpolation is very cheap to compute compared to the whole multiresolution filtering. The original and the modified version of a motion produced with multiresolution filtering have very predictable differences compared to real performances by actors. This predictability makes it possible that, if the original and the modified motions are both plausible, then all the interpolated motions are also unproblematic. If an actor performs two different versions of an action, there are no guarantees that the actions can be interpolated without causing visible problems.

Changing the length of the motion paths did produce good looking results for many different motions. Most of the problems were related to exaggerating motions that are already near the limit of what is possible. For example if a boxer makes a punch forward that stretches the arm totally straight, what should happen if the motion is exaggerated? This was tried out and the result was that the trajectory of the hand was not straight anymore. The trajectory was longer and more curved. This transformed the original decisive punch to look like a much more desperate blow.

3.4.2 Posture

Many actions can be done in a similar way even if the limbs and the head are in different positions. Quite small changes in the posture can affect the emotion that we see in the character. Examples of the poses can be seen in the pictures of figure 21.

In order to make variations to posture, we first need to capture the action that we want to modify. The action must be performed in a relatively *neutral* emotion. This way it can be assumed that the posture of the character is near the regular posture of the actor. Then we can capture the still poses that we want to use in the modification. One of the poses must be similar to the *neutral* posture. Figure 21a shows the *neutral* pose of a male actor. Once we have the *neutral* pose we can also capture poses that are variations of the original one. These poses are shown in figures 21 b-e.

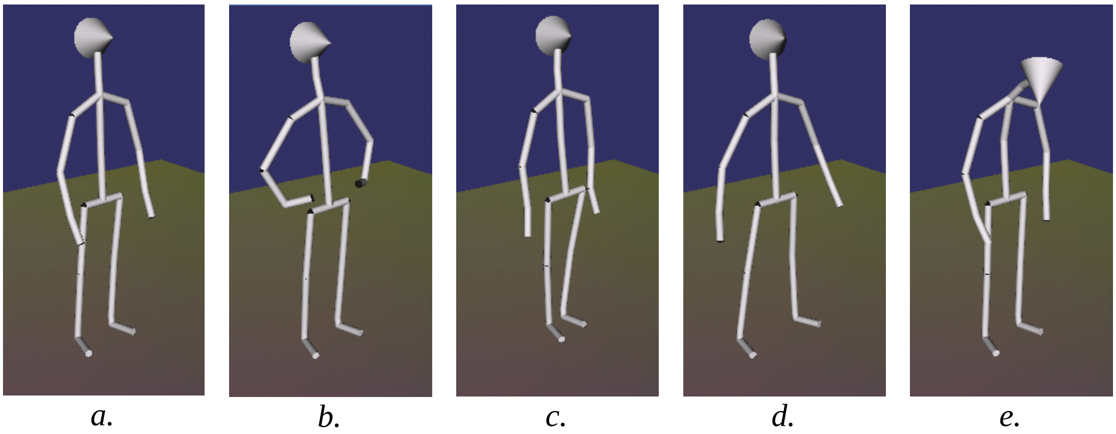


Figure 21: Successfully used still poses in changing the posture of the animated character. Pose a. is the neutral one and poses b. to e. are considered variations of the pose a.

When we have all the poses we can calculate the differences between joint rotations in the *neutral* pose and the other poses. Then we can simply add the desired difference to each frame of the action to change the posture of the animated character. Since the additional operation is only a simple sum for each joint, the computational cost is very low. If we need to reduce or exaggerate the change of the posture, we can multiply all the changes with a constant. The change of the posture was tried out with joint angles and joint directions and both formats worked equally well.

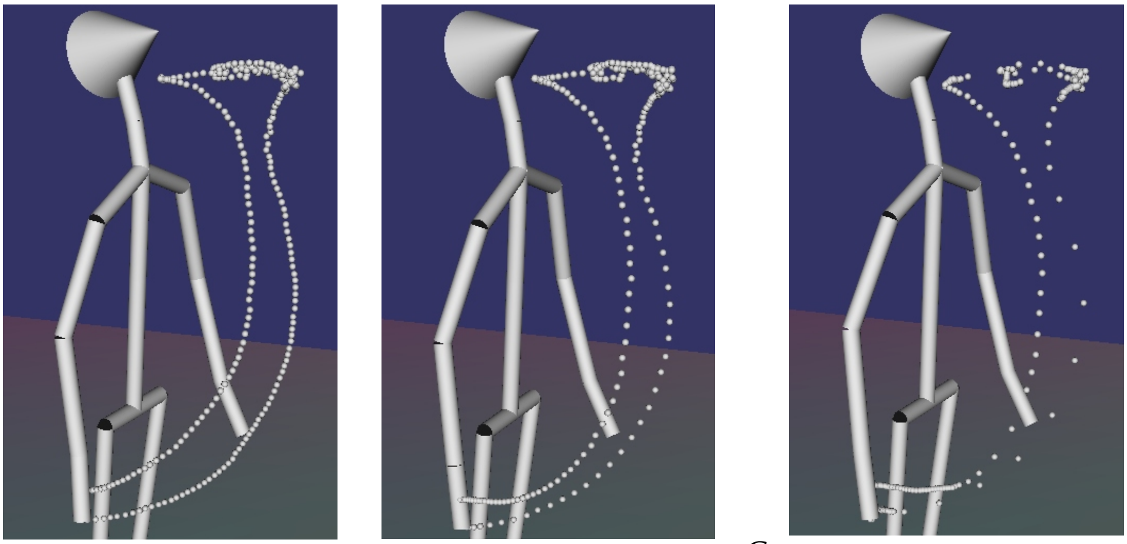
With a motion like a regular walk the change of posture works quite well as it is. If the motion has parts where limbs go very far from their positions in the *neutral* pose, the technique can have unwanted side effects. For example if the character is reaching forward with a straight arm, the arm can become twisted by the change of the posture. This can easily be fixed by gradually fading out the change if the end of the limb goes too far from its position in the *neutral* pose. The fade out must be done to all the joints that affect the position of the limb. For a hand this would include collar, shoulder, elbow and wrist joints.

Capturing the poses must be very well controlled. Otherwise the poses might end up being unusable. The poses must be slight changes from the original *neutral* standing position. That means that it is best to advice the actor to return to the original pose between the changed poses. The hands must be on the side of the body, otherwise they might go through the body when moved. The legs can be moved sideways and the head can move quite freely. In all poses the balance of the actor must remain quite similar and best results were produced with nearly symmetrical poses. If the pose leans to one side, the result can be a character that looks like it is falling down instead of standing.

While capturing the poses was more difficult than was thought earlier, transferring the changes in the posture from one character to another was quite easy. For all the characters the basic *neutral* standing pose is needed. When a set of working differences between the poses was calculated, it was possible to copy the differences to another character without any changes. This worked without any major problems even when the actors were of very different sizes.

3.4.3 Speed and acceleration

A lot of ideas for the implemented modification adopted from the speed transformation of Amaya et al. (1996). Their approach requires comparison of two motions, therefore it could not be used straight away. Still a working modification of speed for a single motion did eventually have a lot in common with their technique.



A. *B.* *C.*

Figure 22: The trajectories of the right hand during a knocking motion. In the middle is the original motion. To the left is the same motion with the speed of the hand made nearly constant. To the right is the motion with added acceleration. The dots in the pictures are the positions of the hand in each frame of the animation.

The idea behind the modification of speed and acceleration is that motions can be divided into two sets. In one end of the spectrum are motions that have a constant speed and almost no acceleration. The other end has motions that are stopped in certain poses and then accelerate heavily when going from one pose to the next. Examples of these motion types can be seen in figure 22.

The modification is done to each limb separately and all the joints of the limb are modified in the same way. For all the limbs the speed is measured from the end point of the limb. That is usually the part that moves fastest, but it is possible to imagine exceptions. For example if you hold your hand near your shoulder, the most moving part of the limb would be the elbow.

The speeds of limbs were measured by tracking the positions of the feet, hands and the head relative to the position of the hips. This way the velocities of limbs do not get distorted for example when the actor is walking. The speed was calculated as difference in positions between the adjacent frames. The resulting signals had a lot of high frequency noise (over 12.5 Hz) that was smoothed out with a low pass filter.

The direction of the motion was not taken into account, because it was estimated that the motion capture system was precise enough to detect even small changes in the amount of the speed. Real motion that is done by actors should follow the laws of physics. It means that sudden changes in the direction of speed should not be possible as that would require very large forces affecting the body. This does not apply to all computer generated motion, therefore the modification might not work correctly with computer generated motion.

Our next step is to calculate how we should stretch the flow of the time for the limbs to get the desired effect. For one frame the relationship between the speed of the limb (v), the time the frame is shown (t) and the distance the end of the limb travels during the frame (s) is:

$$v=\frac{s}{t} \tag{4}$$

To make the speed constant we can substitute the time (t) with the measured difference in positions between the adjacent frames (s) as follows:

$$v=\frac{s}{t},\ \ t=s\ \ \Rightarrow\ \ v=1 \tag{5}$$

With this substitution we have defined an animation that has varying timing of frames. We can return to the original evenly timed format by interpolating new positions for the frames. Now we have an animation where the limbs move at constant speed at all times, but the result does not look quite natural as there are no pauses in the motion. The original pauses are lost, because with the earlier substitution the paused frames are displayed for only a very short time. To fix this issue we have to find the original pauses where the speed of the limb is nearly zero and make those frames last longer. The fix makes the speed slow down from time to time as can be seen in figure 23. Pseudocode for making the speed constant is the following:

```
For each limb do
        Measure speed as the difference between the adjacent frames of the end
        of the limb;
        Smooth out the measured speed with low pass filtering;
        For each joint in the limb do                                          (6)
                Change the frame lengths to be equal to the value of the speed at
                the frame;
                Make the frames that were originally pauses in the motion last
                longer;
                Interpolate evenly spaced values from the signal with varying
                frame lengths;
        Done;
Done;
```

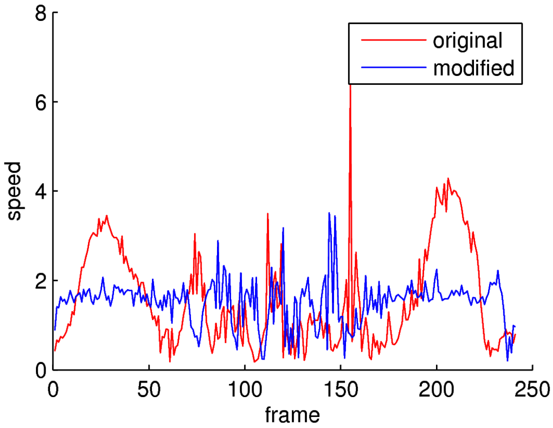


Figure 23: The red curve is the original speed of left hand in knocking motion and the blue is the modified version where speed has been made more even.

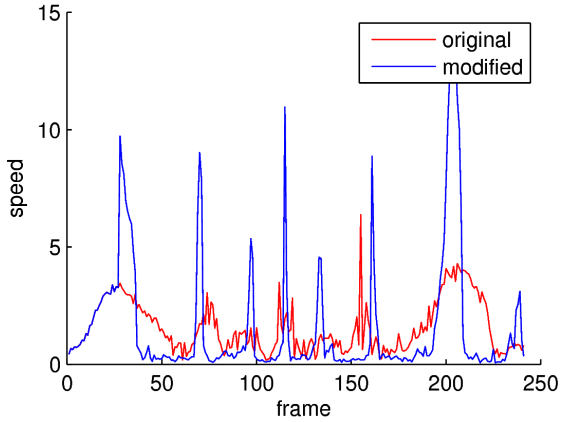


Figure 24: The red curve is the same original speed of the left hand as is in figure 23. The blue curve is the modified version with added acceleration. The modified version of the motion has more slow parts and high acceleration between the slow parts.

When adding acceleration to a motion, the speed of each limb is measured and smoothed in the same way as earlier. Then we have to think about the animation as an ordered set of frames that each last 10 milliseconds. Instead of changing the parts where the acceleration is high, we have to find out the pauses in between the fast parts and make the paused frames last ten times longer than originally. It was found by trial and error that allowing only one pause to be detected per 15 frames produced good looking results. Also, the parts of the motion where the speed stays very low longer that 15 frames should be left without any changes.

The changes increase the total time of the motion, but the original length is returned when the evenly timed frames are interpolated. Since the parts with a lot of acceleration are have not been modified they become faster than before. The resulting change in the flow of time can be seen in figure 25 and the resulting change in the signal in figure 24 . The pseudocode for adding acceleration to a motion is the following:

```
For each limb do
        Measure speed as the difference between the adjacent frames of the end
        of the limb;
        Smooth out the measured speed with low pass filtering;
        Create an array with the duration of every frame;
        Divide all the frames into sets of 15 frames;
        For each frame set do
                If the set is a local minimum of speed and the maximum speed in
                the set is considerably greater than zero
                        Increase the duration of the frames ten times;
        Done;
        For each joint in the limb do                                        (7)
                Interpolate evenly spaced values from the signal with varying frame
                lengths while returning the whole motion to original length;
        Done;
Done;
```

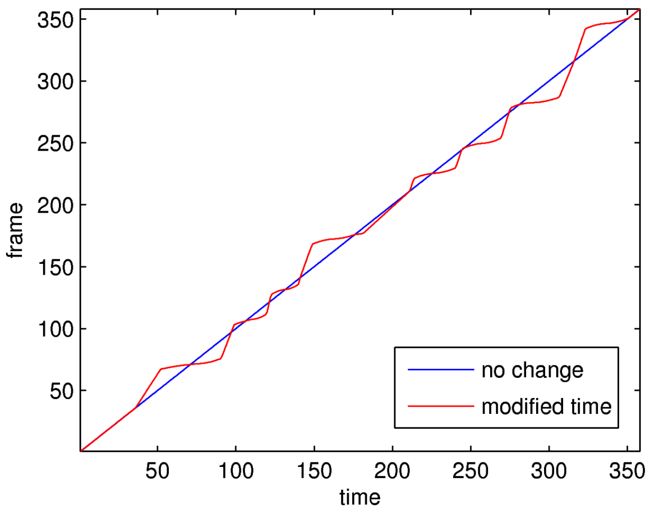


Figure 25: The blue line shows the original relationship between the animation time and the frames of the animation. The red curve shows how the relationship is changed by modification to add acceleration.

It is possible that the modification of speed and acceleration causes phase shift between the limbs. This is often not a problem, but it can look weird. To prevent too much phase shift between the limbs, the motion of the limbs must be processed in short segments. The length of the segment should stay unchanged while the timings of the frames inside it change. When making the speed constant, the segments can be as long as one and half seconds. When adding acceleration to the motion, shorter segments are needed. The best results were produced with segments that were less than a second long. The phase shift that remains is visible in figures 24 and 23. If there would not be any phase shift, the spikes in the speed of the original motion and the modified one would be in the same positions. With the segmentation, the phase shift is not too large to prevent linear interpolation between the original and modified motions. Still, the results of interpolation are not as good as when interpolating motions with modified motion path lengths.

The modification of speed and acceleration works well with motions that can naturally be divided into parts. For example a punch can easily be transformed with the modification to look like slow and weak or fast and powerful. When a motion is very continuous the modification has little effect on it. For example a waving of a hand in a circle might be totally unaffected by the modification, because the speed of the hand could be constant. The main problems of the modification come from the phase shift that it can add between the limbs. For example when walking the legs work in a very synchronized way and the modification can break this link. For these problematic motions it was found easier to revert to the original motion instead of trying to fix the problem.

When the modification of the speed is done by stretching the time similarly for all the joints of the limbs, the format of the joint rotations has no effect on the results. The path that the end of the limb does relative to the start of the limb is unchanged by the modification. Only the speed along that path is changed. We made also a few tests with stretching the time for all of the joints independently, but the results were not promising. The phase shifts between the joints made the motions look noisier. The phase shift also caused random looking changes in the motion paths. This approach was abandoned, because of the bad results. We made also one test that tried to change a motion by making the joint signals straight lines between the minima and the maxima. This was found to have only very little effect on how the motion looks, therefore the approach was also abandoned.

3.4.4 Other modifications

Amplifying high frequencies with multiresolution filtering to add shaking to the character did not produce good results and seemed only to add noise to the motion. Another way to simulate nervous shaking is to add generated or recorded noise to the signal. Both ways to add shaking were tried out and both of them produced results that looked unnatural. One problem was that the hierarchical skeleton amplified the shaking, making small movements in the parent bones cause large movements in child bones. Another problem was that shaking can easily make the feet slide on the floor when they should remain still. Fixing these problems was considered too much work, therefore adding noise to the signal was abandoned.

Changing the speed and acceleration for the limbs did produce usable results, but only for a limb at a time. Changing the speed for the whole skeleton was also tried out. It was done simply by changing the timing between the frames. This made the character look as it was in slow motion or in fast forward. The approach was abandoned, because effect was not considered helpful for changing the emotional content.

Shapiro et al. (2006) used independent component analysis (ICA) to divide the motion of a character to components. Since multiresolution filtering also divides the motion signal to components the ICA seemed like a potential alternative way to make the division to components. When experimenting with the ICA it became apparent that it cannot be used in the same way as multiresolution filtering. ICA is not a stable and predictable algorithm and it also requires a human to examine the components in order to be used in a meaningful way.

3.4.5 Summary of the modifications

The following figure 26 holds short summary of the implemented modifications, their aims and the results of the implementation. The results state if the implementation worked enough well to be useful in animations.

Modification	Aim	Result
Length of motion paths	To exaggerate and diminish the range of a motion.	Works. Results are not always physically accurate. Requires post-processing to fix foot plants. Used in the questionnaire.
Posture	To change the posture of the character.	Works. Requires very little computing. Can make a character appear unbalanced. Used in the questionnaire.
Speed and acceleration	To change a motion to constant speed or to add acceleration.	Works. Can make feet go out of sync. Used in the questionnaire.
Adding recorded or generated noise to motion	To add nervousness to character.	Did not work. Makes the motion unnatural.
Changing the overall speed of animation	To make motion faster or slower.	Works. Gives the look of slow motion or fast forward.
Independent component analysis	To create a mix two motion styles with the system.	Works. The algorithm is not stable and the mixing requires a human operator.

Figure 26*: Summary of the modification to motion data.*

3.5 Repairing modified motions

In previous parts of the text, three feasible ways to modify motion capture data were presented. These three ways do produce sensible changes in the motion, but there are no guarantees that the motion remains physically plausible and free of visible artifacts. To ensure the quality of the motion post-processing is necessary. In many cases even quite simple post-processing can help a lot.

In most cases detecting problematic parts of motions can be easily done by viewing the animation, but it of course takes a lot of time and is prone to human mistakes. That is why automatically detecting problems can be very advantageous. We can fix the problems by removing the modification that caused them or we can try to modify the motion even more in a way that makes the motion look natural again.

When modifying the posture of the character, most of the errors appear when the limbs are far away or very close to the torso. The former situations are already taken into account as the change in the posture is faded out at those times. When the limbs are close to the torso there is a risk of an intersection. Finding them would require geometric collision detection. Since this time only a stick figure was used, detecting the intersections of body parts was not done.

Making changes to the motion paths can cause at least three types of problems. First one is that the feet can start sliding on the floor. The sliding feet are not a problem if the character is standing still, but in all other cases the feet are very likely to slide. Second type is that the joints can go out of their normal range. Third type is that body parts can intersect each other in the same way as mentioned before.

To fix the foot positions a very simple approach was taken. The original foot plants were detected by measuring the distance of the feet from the floor and the speed of the feet. A foot is planted when the distance and the speed are both zero. Fixing the sliding at those times in the modified motions was done by moving the hips into the opposite direction than the foot moves. The times, when both feet are in the air, were quite short. Therefore it was assumed that the hips continue to move at the same speed as before at those times. This way of fixing the sliding feet does not always work, but most of the time the results are surprisingly good.

Joints going out of their normal range were a visible problem when modifying the motion path lengths with angle based joints. The problem became much less visible when joint directions were used instead of the angles. We tried to to solve the problems with hard and soft limits similarly to Bruderlin and Williams (1995). When trying the limits with joint angles, we noticed that sometimes gimbal lock makes defining the limits impossible. Defining the limits was not a problem when using the joint directions. However, the resulting motions did not look always natural. We deduced that the problems appear, because the limits interfere with the conservation of momentum. For example if the elbow joint locks in a motion, the momentum in the movement should not vanish, but it should be transferred to the movement of the other parts of the arm. Simulating this would require physical modeling and would have required more work than was justified by the relatively small problems in the joints.

Modifying the speed and acceleration of motions can add phase shift between the limbs. Any suitable way to fix the problem was not found. If problem was visible when watching the motions, the only solution was to disable the modification from the problematic limbs.

All the modifications can cause problems in some situations, but often removing only one modification out of the three is enough to fix the problem. There are situations where any modification would break the motion. For example when reaching for an object in a tight space, changes in the motion may cause collisions. In these situations, the only choice is to detect the problems and revert to the original motion.

3.6 Combining motion modification methods

Since three different modifications and post-processing are used, it is important to define how they are used together. During the process conversions to format of the motion data is also needed. In the beginning the motions are defined with angles. All the modification are done in the direction vector format. Lastly the parts that show the motions require them to be defined with quaternions. The following list describes the steps in the process and the tools we used:

1. Capture the *neutral* motion and the different standing poses. (Optitrack Full Body Motion Capture system)
2. Cut the motion to appropriate length. (a custom C++ program)
3. Convert the motion from joint angles to joint directions. (a custom C++ program)
4. Create two new versions of the motion with longer and shorter motion paths. (Matlab script)
5. Fix the sliding feet in the new motions. (a custom C++ program)
6. Save the positions of the ends of the limbs in a file for each motion. (a custom C++ program)
7. Create two new motions, one with constant speed and another with added acceleration from each of the three motions. (Matlab script)
8. Find suitable standing poses from the captured material and convert them into joint direction format. (a custom C++ program)
9. Calculate the differences between the *neutral* and the changed standing poses. (Matlab script)

10. Use the nine motions to interpolate the final motion and add changes to the posture. (a custom C++ program)

C++ programs with OpenSceneGraph library were used because some things were found easier to program using the OpenSceneGraph library than with Matlab. Matlab scripts were used for parts because of the good signal processing and analysis tools that Matlab provides.

To be able to modify path lengths, speed and acceleration at the same time, we must first generate all possible combinations as seen in figure 27. The modification to the path lengths is best to be done before modifying the speed, because this way we only need to fix the sliding feet in two motions. The modification to the posture is left as the last thing to do, because it is a very cheap operation and can be calculated in real time unlike the other modifications that need off-line processing. These operations in the last step are all very fast, giving instant feedback to the animator controlling the system. The animator can then decide how strong changes are the best for the path lengths, the speed and the posture.

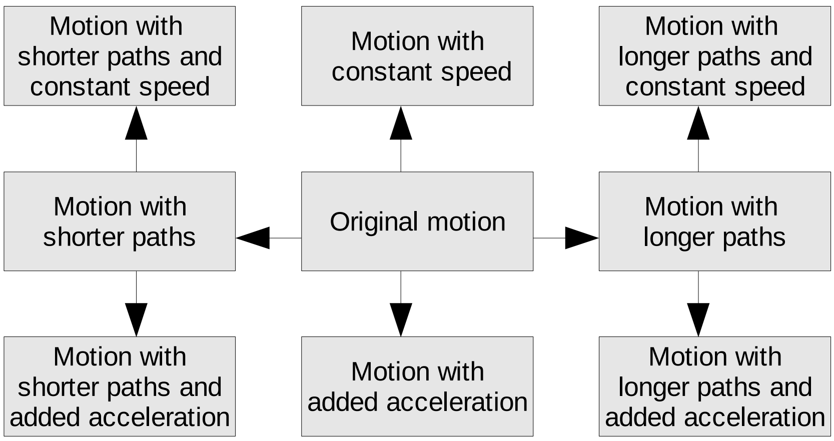


Figure 27: The modifications to the motion paths and the speed and the acceleration create a set of nine motions.

The interpolation between the modified motions is done with two variables that both can have values between 0 and 2. This can be seen as a two dimensional plane as the figure 28. One of the variables affects the length of the motion paths and the other affects the modification of speed and acceleration. The interpolation is done as a weighted average between four motions for each frame. The direction vectors are used, because the quaternions would require a more advanced method of interpolation and angles can have problems caused by gimbal lock.

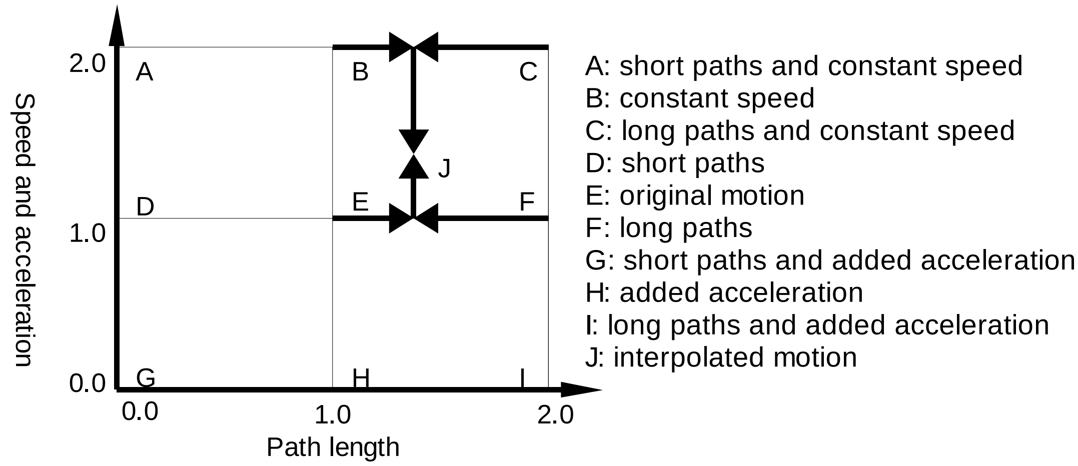


Figure 28: Example process: Motion J can be interpolated from the motions B,C, E and F. The order of the interpolations could be changed without affecting the result.

The described way to combine the modifications only creates new versions of a single motion. This was enough for our evaluation test where only one motion was displayed at a time. However, a version of the process that could modify a motion graph was also created. A motion graph is a set of motions that can seamlessly follow each other. The motion graph that was tested had standing, walking and turning motions. All the motions in the graph were put through the modification process individually. The program that was used to display the motion graph allowed the user to control the movements of the walking character interactively and change the parameters of the modifications in real time. The result was a walking character whose mood could be changed in many ways and that was moved around in the same manner as in many computer games. The modified versions of the motions in the graph fitted surprisingly well together. Even without any additional smoothing the transitions from one modified motion to another with the same parameters looked almost as good as the transitions between the original motions in the graph.

4 Estimating emotional content

This chapter is about testing the emotion content of the modified and acted videos. First, the research questions are presented. Next how the required motions were captured and created is explained. Then the actual questionnaire and the ways it can answer the questions are described. As the last thing, the results that the questionnaire gave are presented.

4.1 Research questions

No thorough testing of emotional content in motion capture data was found in the literature survey. This left many large questions unanswered. It was not even sure that affecting the emotional content of motion captured data was possible to accomplish reliably enough to be useful. This lead the research questions to be quite numerous and of exploratory nature:

1. Can all included emotions be seen in a motion capture animation?
2. Are emotions that are acted or created with modifications seen as the same emotion that was intended?
3. Do the modifications create the intended emotion reliably?
4. How well gender can be seen from motion capture animation?
5. Can the gender that is seen in animation be changed without changing the bone lengths of the character?
6. Are some emotions always seen together in a motion capture animation?
7. Are there any visible side effects caused by the modifications?

Part of the questions are too large to be answered very precisely. Still, it should be possible to find limits to what is possible and what is not. The first question is important, because if an emotion cannot be seen in motion capture animation, it is useless to try to create it with modifications. The second question is relevant, because people might not interpret the emotions in the same way as an actor or an animator. The third question about the reliability of the modifications is important, because unreliably modifications are not useful. The questions related to the gender of the character in the animation are interesting, because gender can have an effect on the emotions we see. The sixth question is useful for future questionnaires, because it does not make sense to give too similar choices in a questionnaire. The last question is relevant if the modifications are to be used in a realistic animation.

4.2 Capturing and creating motions with varying emotions

In order to test if the modified motions actually change the emotional content, we captured motions with different emotional states with actors performing a few steps of walking and a knocking motion. The set of emotions was decided to be *afraid*, *angry*, *excited*, *happy*, *neutral*, *relaxed*, *sad*, *strong*, *tired* and *weak*. These emotions were considered familiar and sufficiently and different from each other.

One of the actors was female and the other a male. Both of the actors had acted in theater performances, but they did not have previous experience with motion capture. The performances were captured in two separate sessions with only one actor present in a session. The actors were familiarized with the principles of the motion capture system and showed what kind of data the system captures. They were also instructed to exaggerate all the emotions. It was pointed out that the motions were the only captured part of the performance, therefore any facial expressions would not show in the resulting animation. It was told the actors to avoid any gestures unrelated to the walking and knocking. The emotions were named to the actors in English, but other instructions were given in Finnish language.

We also captured many standing poses with both actors. These ended up being mostly useless as the poses were very often too unbalanced when compared to the *neutral* standing pose. Luckily the changes in the postures were found out to be quite easily transferable from character to another, therefore we did get enough poses to make the modifications.

It was obvious after viewing modified motions that emotions *sad* and *strong* could be created with changes to posture. The *masculine* and *feminine* postures were also found in the same way. *Angriness* was decided to be created with increased acceleration in the similarly as Amaya et al. (1996) had done. The combination of *sad* posture and long motion paths were considered interesting, because the combination seemed to also result in motions that looked *angry*. Reduced acceleration was used to create *relaxedness* as *relaxed* can be interpreted as opposite of *angry*. Selecting the modifications for other emotions required viewing the unmodified emotional performances and selecting the modification based on them. The resulting combinations of the modifications are in the figure 29.

Emotion	Modification
afraid	Shorter motion paths, hands a little closer to body and feet a little closer to each other.
angry	Longer motion paths and increased acceleration.
energetic sadness	Longer motion paths and head looking down.
excited	Longer motion paths and a little more acceleration.
feminine	Hands closer to body and feet closer to each other.
happy	No modification.
masculine	Hands farther from body and feet a little more apart.
relaxed	Reduced acceleration.
sad	Shorter motion paths and head looking down.
strong	Longer motion paths and elbows away from the body.
tired	Very short motion paths and head looking a little bit down.
weak	Very short motion paths, hands a little closer to body and feet a little closer to each other.

Figure 29: Desired emotions and the modification done to achieve them.

4.3 Questionnaire

To get answers to the research questions a web based questionnaire with 40 different videos was made. Half of the videos were unmodified motions that were performed by an actor trying to make the motion in one of ten emotional states. The emotions were *afraid*, *angry*, *excited*, *happy*, *neutral*, *relaxed*, *sad*, *strong*, *tired* and *weak*. One video of each emotion with no apparent problems was chosen from the performances made by both female and male actor.

The videos with modifications were made using the male and female *neutral* motion as the starting point. For each emotional video performed by an actor a modified version was made that tried to show the same emotion as the actor. This resulted in 16 videos as the *neutral* and *happy* videos did not get a pair. The modified versions of the videos with a *happy* emotion were left out, because no effective modification was found to increase the *happiness* of the animation with the modifications. The list of all the videos is in appendix 1.

Four special videos were also included in the questionnaire. For the male actor one video was made that tried to make the posture of the character more *feminine* and another video that combined the modification for *sad* posture and long motion path lengths. For the female actor one video was made that attempted to make the character more *feminine* and the same modification reversed was used to make the character more *masculine*.

It might be hard to evaluate a video without seeing any other videos for comparison. For this reason there were two randomly selected videos shown to each participant before the actual 40 videos. Answers to these two videos were not used in the analysis of the results, but this was not told to the participants. All of the 40 videos were shown in randomized order. The participants were able to play the videos many times, but it was instructed that viewing the video once should be usually enough.

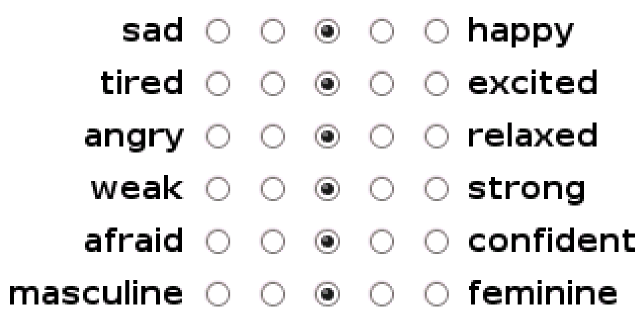


Figure 30: The choices that the participants were asked to make for each video in the questionnaire.

To make the questionnaire simple enough to answer, it was decided to pair opposite emotions. The pairs were *sad-happy*, *tired-excited*, *angry-relaxed*, *weak-strong*, *afraid-confident* and *masculine-feminine*. The emotion *confident* was not part of the original set of emotions, but it was necessary to have a pair to the emotion *afraid*. For each video, the participants were asked to evaluate which of the adjectives describe the character in the video best in a scale with five steps as is seen in figure 30. The closest choice to the adjective was instructed to be used if the adjective describes the character very well. The middle choice was instructed to be used if neither of the alternatives feel good or if the participant is unsure which one is better. The default choice for all selections was the middle one. The questionnaire also included a text box for comments. The comments were instructed to be given if something special caught attention in the video.

The questionnaire was made with a server side PHP script. The videos were embedded to the web page as Flash objects. Internet Relay Chat (IRC), newsgroups and social media were used to get participants for the questionnaire. The participants did not receive any compensation for answering. There were 28 participants that answered the whole questionnaire. 8 of the participants were female and 20 were male. Answers were mostly given with the selections. The free text answers were given only 29 times while the participants had the chance to make comments 1176 times.

4.4 Results

To analyze the answers of the questionnaire, the answers were interpreted as numbers from -2 to 2. This way for example very *sad* would be -2, a little *sad* -1, a little *happy* +1, very *happy* +2 and 0 would be *neutral* between *happy* and *sad*. From these numbers a distribution can be calculated for each emotional axis of a single video or of a set of videos. The distributions can be assumed to be enough similar to normal distributions, therefore t-tests can be used to analyze the statistical significance of the answers.

The first research question asked, if all the included emotions can be seen in a motion capture animation. To answer the question for one emotion, we should compare the video with the highest scores for the emotion and a comparison group of all the videos that were acted by the same actor. If the emotion was really seen by the participants, the answers for the emotion in the highest scoring video should create a distribution that is significantly different from the distribution created by the answers of comparison group. The distributions were compared by testing the equality of their means with t-tests and the result are shown in figure 31. Result '0' from the t-test tells that the distributions have equal means and the result '1' indicates that the distributions have significantly different means.

Emotion	Result of t-test, 0=equal, 1=different	p-value	Difference between means	Gender of the actor	Was the video acted or modified from neutral motion	Emotion intended by actor or animator
relaxed	1	0,002	0,62	male	modified	excited
angry	1	p < 0,001	0,91	female	acted	angry
happy	1	p < 0,001	0,98	male	acted	excited
masculine	1	p < 0,001	1,02	male	modified	strong
feminine	1	p < 0,001	1,13	female	modified	womanly
confident	1	p < 0,001	1,17	male	modified	strong
strong	1	p < 0,001	1,3	male	modified	strong
excited	1	p < 0,001	1,31	female	acted	excited
sad	1	p < 0,001	1,4	female	modified	sad
tired	1	p < 0,001	1,55	female	acted	tired
weak	1	p < 0,001	1,61	female	modified	tired
afraid	1	p < 0,001	1,71	female	modified	tired

Figure 31: Results of the t-test with 5% significance level comparing the means of answers in the same axis as the emotion. The compared videos are the ones with the highest mean for the emotion and the set of all videos of the same actor.

T-tests of the figure 31 show that, in the best case, every emotion did create a distribution that was significantly different from the distribution of the comparison group. This means that all of the included emotions were seen by the participants in at least in one video.

Figure 31 also has information about the amount of the difference between the means of the answers, gender of the actor in the compared videos, the type of the motion, if the video was acted or modified and the emotion that was intended by the actor or the animator. Since the scale for all of the axes goes from -2 to +2 the maximum possible difference between the means is 4. Comparing the listed emotions and the intended emotions shows that the highest scores did not always go to the intended video. To visualize the data in figure 31, the distributions of the answers of the best scoring videos are shown in figure 32. The means of all the videos can be seen in appendix 3.

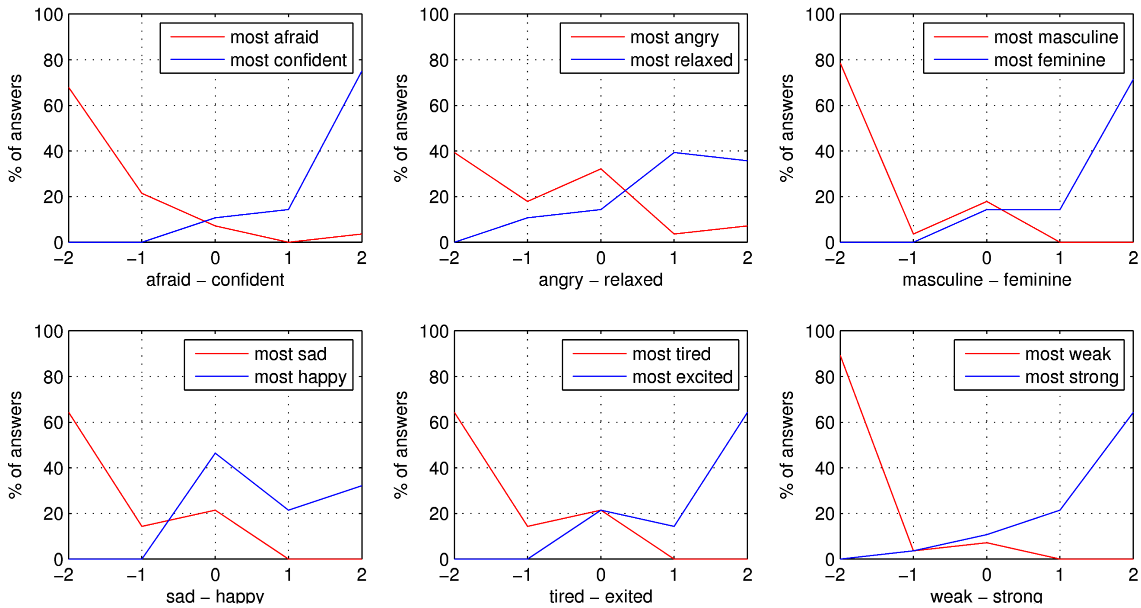


Figure 32: The distributions of answers of the ten videos that were rated most emotional. The same videos that were used in the t-tests of figure 31.

The second research question asked if emotions that are acted or created with modifications are seen as the same emotion that was intended. If the intended emotion would be perfectly identified by the participants, the intended emotion would get a larger percentage of the scores than any other emotion.

Figures 33 and 34 are confusion matrices for the acted and modified videos that show the percentage of the scores that each emotion was given. Each of the perceived characteristics was allowed to get zero or one or two points from one answer. For example if an answer considered a video to be very *masculine* the video got two points and if it was considered a little *masculine* the video got one point. If the video was *feminine* instead of the *masculine,* the point would go to *femininity* and the *masculinity* would get zero points. These points would then be changed to percentage by dividing by the total points of the row. All of the rows, except for the energetic *sad* in the modified videos, have the points of the male and the female videos counted together. The energetic *sad* modification was done only to the motion of the male actor. The percentage is underlined in the matrices if it was the intended emotion.

Intended emotions %	Perceived characteristics: weak	tired	afraid	sad	relaxed	angry	excited	happy	strong	confident	feminine	masculine
weak	20	12,16	20	20,78	3,53	3,14	1,57	1,18	1,96	2,75	5,1	7,84
tired	16,27	28,17	4,37	12,7	8,33	4,76	0	0,79	4,76	7,14	0,79	11,9
afraid	21,79	5,98	31,2	8,12	3,42	1,28	4,7	0,85	2,99	3,42	11,97	4,27
sad	15,25	18,39	13,9	17,94	8,97	2,24	0,45	0	3,14	4,93	4,93	9,87
relaxed	8,56	9,46	3,6	2,7	13,96	1,8	3,6	3,15	11,71	20,72	8,56	12,16
angry	1,78	0,36	0	0,71	8,19	11,39	15,66	4,98	16,73	27,05	2,14	11,03
excited	1,32	1,32	0,66	0	9,9	3,63	21,12	14,52	11,22	21,45	8,91	5,94
happy	0,41	1,22	0,41	0	12,24	5,31	13,47	6,53	12,65	26,53	9,8	11,43
strong	2,34	2,73	1,17	1,56	14,84	5,08	10,55	8,59	12,5	23,83	10,94	5,86
neutral	8,47	10,05	6,88	3,17	18,52	1,59	4,23	4,76	7,94	14,29	10,58	9,52

Figure 33: Confusion matrix between the acted videos and the characteristic that were perceived. The intended emotion is underlined and the high percentages are highlighted with orange.

		Perceived characteristics											
		weak	tired	afraid	sad	relaxed	angry	excited	strong	happy	confident	feminine	masculine
Intended emotions %	weak	23,15	18,98	13,43	9,26	6,94	0,93	2,31	0,93	1,85	4,63	12,04	5,56
	tired	28,29	14,8	22,7	14,14	2,96	1,32	1,32	1,32	0,33	1,97	5,59	5,26
	afraid	16,93	17,99	17,46	3,7	7,94	1,06	3,17	4,23	2,12	6,88	12,17	6,35
	sad	20,64	19,57	12,46	28,47	3,2	2,85	0	1,07	0	1,78	2,85	7,12
	relaxed	6,91	8,51	7,45	3,19	13,3	2,66	5,85	10,11	3,72	17,55	10,64	10,11
	angry	1,02	0,34	0,34	0,68	13,61	2,72	11,56	15,31	7,82	23,13	8,84	14,63
	excited	1,38	2,41	0	0,69	19,66	2,07	11,38	13,45	9,66	20,34	8,62	10,34
	strong	0,64	0,32	1,6	0	5,77	6,41	8,65	22,44	5,13	24,68	0,32	24,04
	energetic sadness	6,9	14,37	5,17	20,11	2,87	13,79	2,3	8,62	0	12,07	1,15	12,64

Figure 34: Confusion matrix between the modified videos and the characteristic that were perceived in them. The intended emotion is underlined and the high percentages are highlighted with orange.

The third research question asked if the modifications create the intended emotion reliably. The data from the questionnaire enables three ways of answering the question. The first way is to compare the modified motions with the original *neutral* motion. The modifications should increase the scores of the intended emotion if they are reliable. Figure 35 shows the difference of the means scores between the modified and the original motion for each emotion. A positive number indicates that the modification has increased the amount of the emotion.

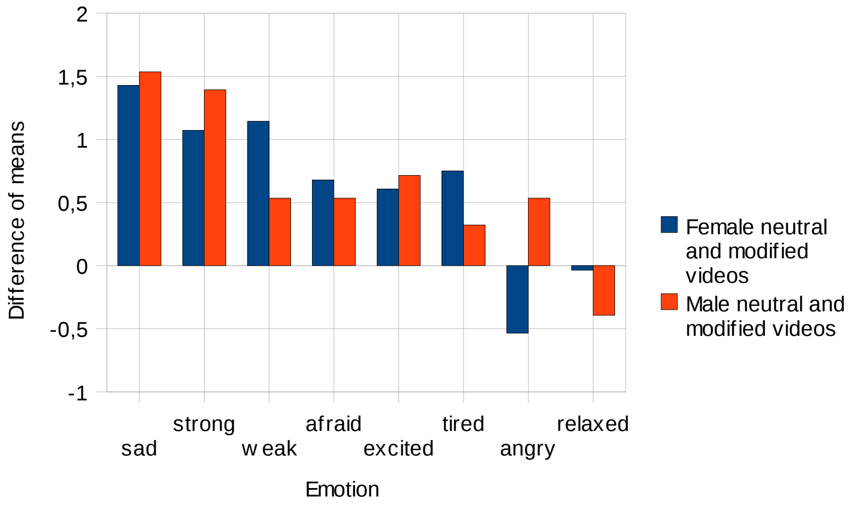


Figure 35: Difference of means of answers between the modified videos and the neutral videos. A positive value means that the modified video has more of the intended emotion than the neutral.

The second way to examine the reliability of the modifications is to compare the modified motions with the motions that got the highest scores in the same emotion. If the modifications are working well, there should not be any videos that score higher points in the same emotion. Figure 36 shows the results of this comparison. The figure has the difference of the means between the modified and the highest scoring videos. Value of zero means that the modified video had the same score as the highest scoring video. The amount of difference is useful when comparing the successfulness of the modifications.

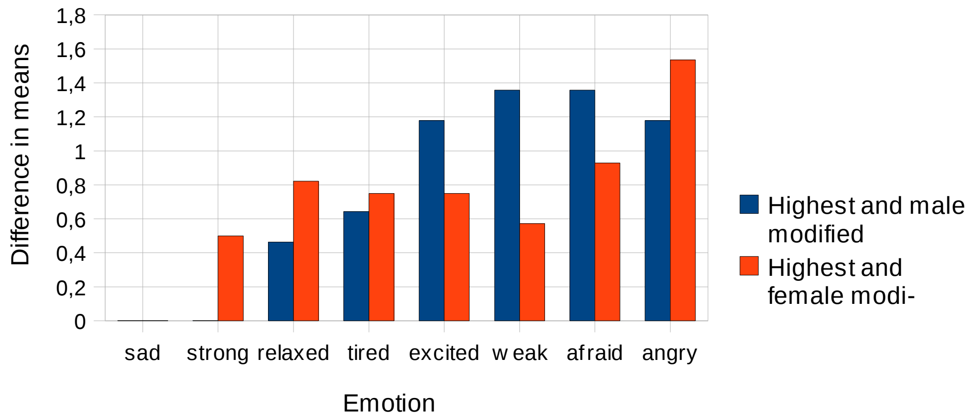


Figure 36: Difference of means of answers between videos that have the highest scores for the emotion (listed in figure 31) and the modified videos that were intended to show the emotion. A low value means that there is only a small difference between the highest rated video and the modified video in the axis where the intended emotion belongs.

The third way of examining the reliability of the modifications is to compare the modified motions with the acted motions that intended to have the same emotion. To get a clear answer which ones were better a comparison of mean scores and t-tests were necessary. The results are shown in figure 37. Since there were only two actors the results cannot be generalized. The t-test was done with significance level of 5%. The amount of answers was not always enough to get a statistically significant difference between the acted and the modified videos. Those rows have been grayed out in the figure.

	Emotion	Result of the t-test, 0=equal, 1=different	Video with highest mean	Difference between means
Male actor	afraid	1	acted	1,11
	angry	0	modified	0,07
	excited	0	acted	0,46
	relaxed	0	-	0
	sad	0	modified	0,36
	strong	1	modified	1
	tired	0	acted	0,32
	weak	0	acted	0,11
Female actor	afraid	0	acted	0,5
	angry	1	acted	1,54
	excited	1	acted	0,75
	relaxed	0	acted	0,25
	sad	1	modified	1,07
	strong	1	modified	0,5
	tired	1	acted	0,75
	weak	0	modified	0,18

Figure 37: Comparison between the means of answers of the acted and the modified videos in the axis where the intended emotion belongs.

The fourth research question was about how well gender can be seen from motion capture animation. For this question a comparison of the videos of the female and male actor was done. The acted videos of the female and the male actor have statistically significant different means in the *masculinity-femininity* axis that is seen in figure 38. The same difference can be seen in the distributions of figure 39. As previously, since there were only two actors the results cannot be generalized.

Result of t-test, 0=equal, 1=different	p-value	Difference between means
1	p < 0,001	1,03

Figure 38: Test between the means of answers of acted videos of man and woman actor in masculine - feminine axis. The t-test has significance level of 5%.

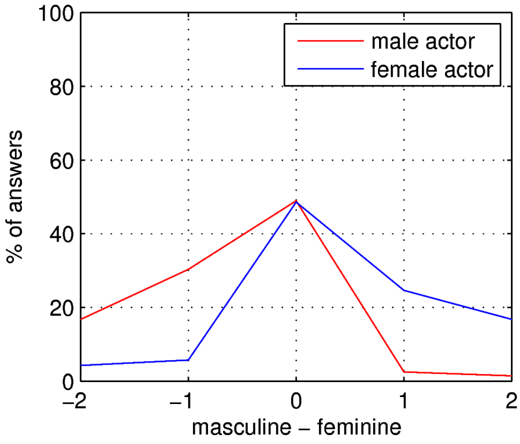


Figure 39: Distribution of answers of acted videos of man and woman actor in masculine - feminine axis.

The fifth research question asked if the gender seen in the animation can be changed without changing the bone lengths of the character. Since all the modifications kept the bone lengths of the skeleton unchanged, a comparison of the *femininity* and *masculinity* of all the modified and acted videos was done. The results are shown in the figure 40.

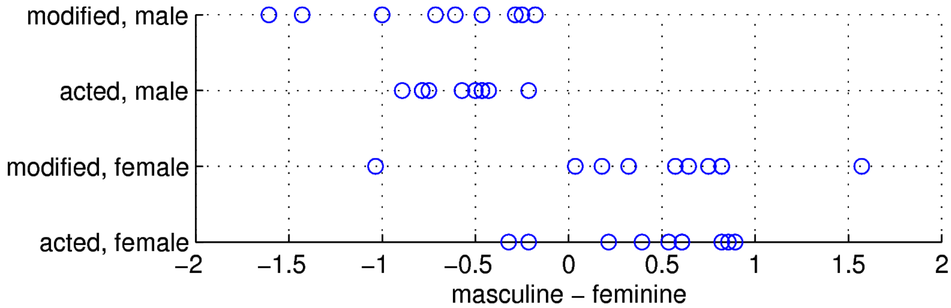


Figure 40: Mean answers in the masculine - feminine axis for the all videos.

The results in the figure 40 show that a few of the modified videos did have more *masculinity* or *femininity* than the acted videos of the same actor. A comparison of the mean scores of the most extreme cases and the original motions they were based on are in the figure 41. The distributions used in the t-test of the figure 41 are visualized in the figures 42 and 43.

	Result of t-test, 0=equal, 1=different	p-value	Difference between means
male, neutral vs. male, modified, feminine	0	0,19	0,29
male, neutral vs. male, modified, strong	1	p < 0,001	1,14
female, neutral vs. female, modified, masculine	0	0,14	0,36
female, neutral vs. female, feminine	1	p < 0,001	1,04
female, neutral vs. female, modified, strong	1	p < 0,001	1,57

Figure 41: Results of the t-tests with 5% significance levels between means in masculine – feminine axis of neutral videos and videos with modifications that had the largest effect on the axis.

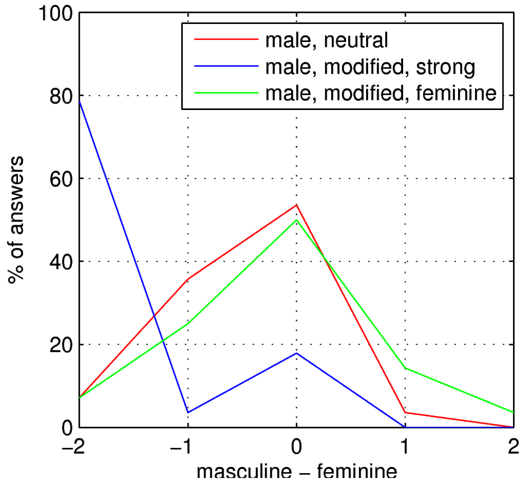


Figure 42: Distributions of answers for neutral video by male actor and the same data modified to look more feminine and stronger.

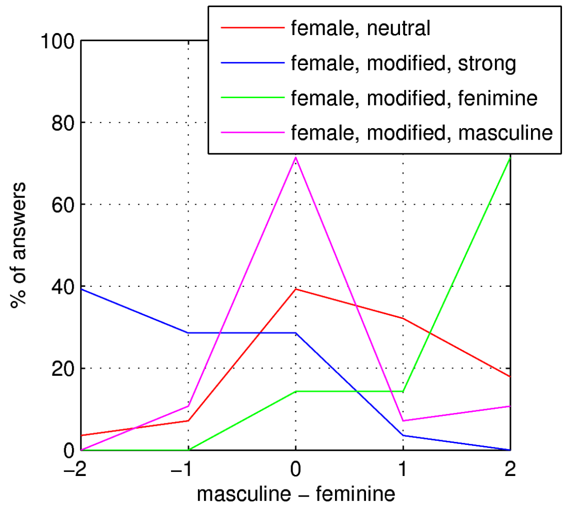


Figure 43: Distributions of answers for neutral video by female actor and the same data modified to look more feminine and more masculine and stronger.

The sixth research question asked if some emotions are always seen together in a motion capture animation. To answer this correlations between the emotions that the participants used when answering were calculated. Since one axis of the questionnaire has two emotions the axes had to be broken up and the results can be seen from figure 44.

	afraid	confident	angry	relaxed	masculine	feminime	sad	happy	tired	excited	weak	strong
afraid	1	-0,46	-0,08	-0,2	-0,16	0,09	0,26	-0,19	0,05	-0,16	0,51	-0,3
p-value	1,000	0,000	0,005	0,000	0,000	0,002	0,000	0,000	0,130	0,000	0,000	0,000
confident	-0,46	1	0,2	0,23	0,31	0,02	-0,24	0,31	-0,25	0,36	-0,44	0,57
p-value	0,000	1,000	0,000	0,000	0,000	0,554	0,000	0,000	0,000	0,000	0,000	0,000
angry	-0,08	0,2	1	-0,22	0,23	-0,1	0,13	-0,08	0,03	0,08	-0,13	0,26
p-value	0,005	0,000	1,000	0,000	0,000	0,001	0,000	0,008	0,312	0,009	0,000	0,000
relaxed	-0,2	0,23	-0,22	1	0,16	0	-0,12	0,38	-0,02	0,13	-0,15	0,16
p-value	0,000	0,000	0,000	1,000	0,000	0,885	0,000	0,000	0,603	0,000	0,000	0,000
masculine	-0,16	0,31	0,23	0,16	1	-0,31	0,02	0,06	0,08	0,06	-0,17	0,41
p-value	0,000	0,000	0,000	0,000	1,000	0,000	0,431	0,058	0,006	0,032	0,000	0,000
feminine	0,09	0,02	-0,1	0	-0,31	1	-0,08	0,11	-0,12	0,14	0,05	-0,02
p-value	0,002	0,554	0,001	0,885	0,000	1,000	0,005	0,000	0,000	0,000	0,086	0,457
sad	0,26	-0,24	0,13	-0,12	0,02	-0,08	1	-0,2	0,41	-0,19	0,32	-0,15
p-value	0,000	0,000	0,000	0,000	0,431	0,005	1,000	0,000	0,000	0,000	0,000	0,000
happy	-0,19	0,31	-0,08	0,38	0,06	0,11	-0,2	1	-0,2	0,51	-0,21	0,23
p-value	0,000	0,000	0,008	0,000	0,058	0,000	0,000	1,000	0,000	0,000	0,000	0,000
tired	0,05	-0,25	0,03	-0,02	0,08	-0,12	0,41	-0,2	1	-0,31	0,31	-0,2
p-value	0,130	0,000	0,312	0,603	0,006	0,000	0,000	0,000	1,000	0,000	0,000	0,000
excited	-0,16	0,36	0,08	0,13	0,06	0,14	-0,19	0,51	-0,31	1	-0,26	0,34
p-value	0,000	0,000	0,009	0,000	0,032	0,000	0,000	0,000	0,000	1,000	0,000	0,000
weak	0,51	-0,44	-0,13	-0,15	-0,17	0,05	0,32	-0,21	0,31	-0,26	1	-0,4
p-value	0,000	0,000	0,000	0,000	0,000	0,086	0,000	0,000	0,000	0,000	1,000	0,000
strong	-0,3	0,57	0,26	0,16	0,41	-0,02	-0,15	0,23	-0,2	0,34	-0,4	1
p-value	0,000	0,000	0,000	0,000	0,000	0,457	0,000	0,000	0,000	0,000	0,000	1,000

Figure 44: Correlations between the emotions in the answers when the axes have been broken apart. Cells where the emotions were originally opposite ends of the same axis are highlighted with cyan. Significant correlations with absolute value over 0,3 are highlighted with light orange. Significant correlations with absolute value over 0,5 are highlighted with strong orange.

The last research question asked if there are any visible side effects caused by the modifications. To answer this the free text answers were analyzed. The small number of the free text answers made it impossible to get a reliable answer to the question. A few answers stated that the character looked drunk, too exaggerated or otherwise funny. Still, anything certain cannot be said as both acted and modified videos had this kind of answers. All free text answers can be found in appendix 2.

5 Discussion

In this chapter the answers to research questions are discussed. Ideas about how to make the modifications and future questionnaires better are also presented.

5.1 Observations from the questionnaire

According to the results in figure 31 it is possible to see all the tested emotions in the animation made from motion captured data. This result is very promising, because it means that a large amount of emotions can be seen from motion data. This does not mean that it would be easy to show a certain emotion when wanted or that only one emotion would be visible in the motion data. It was unexpected that *happy* was also one of the emotions that can be seen from the motion capture data. This might be explained by the correlations seen in figure 44. The *happiness* has a strong positive correlation with excitement. This could mean that *happiness* is seen in the videos that have an *excited* motion. Also the quite poor recognition of anger was unexpected as anger was thought to be very visible as increased acceleration in the motion.

If the intended emotions of the videos would be perceived as the same emotion, the underlined values in figures 33 and 34 would be very large compared to other values. This is clearly not the case. Both figures show that many emotions are usually seen from the same motion. The emotions *weak*, *tired*, *afraid* and *sad* seem to be perceived often together and they are not confused with the emotions *angry*, *excited*, *happy* and *strong*. This can be partly distorted by the fact that the emotions were paired in the questionnaire, therefore it was not even possible to select all of the emotions at the same time. The modified videos that were meant to be *angry* were failures and the modified videos intended to be *excited* were seen more often as *relaxed*. It is also good to note that the videos intended to be *neutral* were seen as quite *relaxed*. In the end, these results tell that more work needs to be done, if we want to create emotions reliably by acting or with modifications to existing data.

The comparisons in figures 36 and 35 are between the modified videos and the original *neutral* videos and the videos which were rated most emotional by the participant of the questionnaire. The figures show that the emotions *sad* and *strong* were quite successfully created with modifications. Both of the modifications made an improvement to the *neutral* motion and had the most emotion when compared to other videos or they were at least very close to the most emotional videos. The modified emotions *tired*, *excited*, *weak* and *afraid* were a little closer to the intended than the *neutral* motion, but did not reach near the amount of emotion that the best videos in those categories did. The modified emotions *angry* and *relaxed* were failures. They could even reduce the amount of the intended emotion seen in the videos. The *angry* modification was a failure also when compared to the video with most anger, but the *relaxed* modification did perform relatively well. This might be explained by looking at perceived characteristics of the *neutral* videos in figure 33. The *neutral* videos were already seen as quite *relaxed*, therefore when the modification made them a little less *relaxed*, the videos might not have lost all the *relaxedness*.

The results in figure 37 show that in this experiment the emotion *strong* was seen better from the modified motions. The emotion *sad* was also better as modified, but only for the male actor. The emotions *afraid*, *angry*, exited and *tired* were seen better from the acted motions, but only for one of the actors. This result does not clearly say if the acted or the modified emotions are more reliable. Also, the results for the actors cannot be generalized to all actors or even the same actors on a different day. The modifications produce the same results every time, but the performance of the actors can vary greatly. This makes it hard to fairly compare the performance of actors to the modifications.

The figures 38 and 39 clearly show that the participants saw a difference in the *masculinity* and the *femininity* between two actors. This might suggest that it is possible to see the gender of the character from the motion, but drawing too general conclusions is not possible based on only two actors. The male actor was quite large and strong looking and the female actor was considerably smaller than the male. An alternative explanation to difference in the *masculinity* and *femininity* could be that it is caused by the differences in size and strength between the actors. It could be interesting to try the same questionnaire with actors of different sizes. This way the correlation between the physical characteristics of the actors and the emotions seen in their motion could be calculated.

The comparison of means in the *masculine – feminine* axis of the acted and the modified videos in figure 40 implies that the emotional content of the videos can affect the seen gender. More details on how the videos performed in the axis can be seen from appendix 3: 3/6. The t-tests in figure 41 confirm that a few of the modifications were able to change the seen gender in a statistically significant way. The performance of the female actor was successfully modified to be more *feminine*. The modifications that intended to make the character *stronger* increased the *masculinity* of both actors. This result is important as it shows that the gender of an animated character can be affected without making any changes in the bone lengths.

There are many statistically significant correlations between the emotions of the questionnaire as seen in figure 44. The three highest correlations are too high to be just coincidental. The pair *afraid* and *weak* is possibly just two words for the same phenomenon. It is apparent if you think about how one could look *afraid*, but not look *weak* at the same time. The pair *strong* and *confident* might be similarly linked, but they might not be exactly the same thing. If *strong* means having a lot of muscles, it might not be the same as *confident*. This is again related to how the viewer understands the words. The pair *happy* and *excited* is not a similar case of words with almost the same meanings. A better explanation for the high correlation could be that by itself the emotion *happy* is very hard to see in motion without a face, therefore the next best way to show *happiness* is to be *excited* and energetic.

There are a few free text answers that indicate that something strange was seen in part of the videos. All the free text answers can be read from the appendix 2. For example there are three videos that have answers saying that the character seems to be drunk. One of these is an unmodified video by an actor, therefore it is not clear if the problems are in the modifications or in using motion capture in general. The fact that even from a stick figure animation the motion can look unnatural suggests that there might be more problems if a realistic human model would be used. This could be worth testing in a future questionnaire.

5.2 Improving the modifications

The emotions *strong* and *sad* did work out quite well with the modifications to the postures. So, there is not much need to make them better. For the other emotions there is definitely a need for improvement. Making small changes to the modifications might help a little, but for significant improvements the approaches must be reevaluated.

The emotions *weak*, *tired* and *afraid* were generated quite well by the short motion paths. When compared to the performances of the actors the modification did not create as much differences between those emotions. This can be seen by comparing the confusion matrices in the figures 33 and 34. More differences might be possible to create with suitable changes to the postures. Also if a model for the muscles of the character would be implemented, it might be possible to make the emotions more convincing. *Afraid* might be expressed by careful motions with stiff muscles, while *weak* and *tired* would be more natural with *relaxed* muscles that let the gravity pull the limbs down.

Relaxedness was not generated well with the modifications. This can be partly caused by the fact that *neutral* motions were already seen as quite *relaxed*. Therefore making them even more *relaxed* might not be easy. The modifications that were meant to give an *excited* impression, were seen as quite *relaxed*. Excitement was supposed to be produced with very long motion paths. This could be a hint that *relaxedness* is not only about the low acceleration of the motions, but about using a lot of space for your motions. This would make the *relaxed* motions similar to *weak* and *tired* motions, in terms of low muscle stiffness, but with more energy in them. Another way of making motions more *relaxed* could be making the time flow slower in the modified motion with long motion paths. This could reduce the forces seen in the character.

The motions were seen as a little more *excited* when their motion paths were made longer. This still did not achieve as good results as the *excited* performance of the female actor. When comparing the modified and acted videos that were both intended to be exited there is a clear difference in the time used per step. Both the male and the female actor have a faster pace in their steps than what is seen in the modified versions. All the three modifications were designed to keep the total time used for an action unchanged, therefore increasing the pace of the steps is not possible with the modifications. A way that might not require a lot of post-processing could be making the motion paths shorter and than making the time flow faster. This way the pace of the steps can be increased while the speed that the character goes forward stays the same. Keeping the total time unchanged was seen necessary, because making the time go faster or slower affects all the forces in the motion. This can be seen for example as increased or decreased gravity. With the results of the questionnaire, it looks like editing the flow of time could be necessary in creating excitement. Making an effort on fixing the unnatural forces that the changed flow of time causes might be justified in future work.

Happiness was not successfully seen when the actors tried to act happily, but the *excited* motions were seen at least a little *happy*. This suggests that *happiness* that is not very energetic is not seen in the motion of the characters. The best that can be done to get passive *happiness* without a face is to avoid the emotions that have a negative correlation with *happiness*. Those emotions include *afraid*, *angry*, *tired*, *sad* and *weak* as can be seen from figure 44.

Anger was not seen in the modifications that tried to create it with adding more acceleration to the motion. Anger was only seen in two of the videos as can be seen from appendix 3: 2/6. Those videos were the acted *angriness* of the female actor and the other was the modified energetic *sadness*. Both of the videos were on the *angry* side, but still a lot of people could not see the *angriness* in them. When watching the acted *angry* performance of the female actor, the walk and the knocking look like they have more acceleration than the *neutral* version of the female performance. The acceleration seems to be largest when the feet touch the ground and the hand touches the door. One factor that could explain the bad results of the *angry* modification is that the modification to speed and acceleration was disabled for the feet. This was done, because walking started to look unnatural with it. This clearly needs to be fixed. Another difference between the modified videos and the acted videos is that the acted versions only have a large acceleration while in coming to contact with another object, and no sudden accelerations if the body can move freely. To do the same with the modifications, all the contact places and times should be recorded. This would allow the modification to only affect those spots. This would either require hardware that could record the contacts or a calculation of the forces in the character and estimation of the contacts causing the forces. The best results might be attained if the new version of acceleration modification would be combined with longer motion paths and a slightly *sad* posture.

The modifications that tried to affect the gender of the character were partially successful as is seen in the figures 41, 42 and 43. The female actor was made more *feminine* by the intended modification to the posture. The modification to make the male character more *feminine* did not work. The strong correlation between the *strength* and the *masculinity* of the characters allowed the *masculinity* to be increased. To make the modification to gender work better, changes to the bone lengths of the characters would probably be needed. Also the walk style itself would probably need fixing. If the character would walk in the same way but had its feet much closer to each other, it would fall over quite quickly in reality. These issues make further modifications to gender quite challenging.

Creating nervousness was tried out by adding noise to the motion signal. That produced unnatural looking motion. A better way to create nervousness could be to randomly vary shorter and longer motion paths during the motion. This way the movement could seem more uncertain. The effect could be also increased by stretching time, making the rhythm uneven in the movement.

All the modifications can create unrealistic movements. In the current implementation, these errors are not detected automatically, but the animator must visually inspect the resulting motions. This is not very efficient and takes a lot of human effort. Detecting problems in the motions would require intersection tests between the body parts, a physics model for finding unnaturally large forces and a system for finding out if the character would fall over. After detecting the problems, it might also be possible to automatically correct at least part of them.

One possible goal for the future might be to make the modifications work in real time with very little delay between the capture and the animation. Changes to posture are easy to do in real time, but other modifications would require a lot of optimizations. The modification to the speed and the acceleration could be made faster if instead of new motions only the jumps in the signals would be defined. This is possible, because the modification to speed and acceleration does not change the shape of the motion paths, but only the current position along the path. Changing the length of the motion paths is currently quite a slow operation and it also requires knowledge of the signal values in the past and the future. Even with very good optimizations or a lot computing power the algorithm would still require a small delay for the separation to frequencies to work correctly.

While modifications to motion can produce visible emotions, it still has a lot of limitations. The final animated character should also include a face and fingers as they can be used to relay emotions. To get a full human model to show all the emotions possible, it would also be necessary to have a large data base of symbolic gestures that cannot be created with small modifications.

5.3 Future questionnaires

The questionnaire gave good data, but the more participants the more reliable the data would be. Two of the people who answered the whole questionnaire did later on say that the questionnaire felt too long and hard to answer. They both said that it was irritating not to be able to see any emotions in many of the videos and if they had not known who had made the questionnaire they would not have completed it. In future questionnaires it might be good to have less videos and less selections per video. Another way to make people more motivated could be to add a few videos that are so exaggerated that they would be entertaining instead of irritating.

What actions are seen in the videos should also be carefully thought out. Walking and knocking are good basic motions, but they have restrictions on how much they can be modified. Walking and knocking movements are meant to manipulate the position of the character and cause a sound. This means that the motions are not primarily for expressing emotions. It might be more efficient to estimate the emotional content of a motion that is primarily for communication. The suitable motion could be for example a character waving hands while explaining something.

The questions in the questionnaire could be made better. Using two adjectives as opposite ends of the axes makes analyzing the results more complicated than measuring only one adjective per axis. Choosing adjectives that can describe the motion accurately is also very important and at the same time the number of the axes should not become too large. Possible alternative ways to evaluate emotional content of motions could be letting the participants decide the meaning of the axes themselves or mapping all emotions to a small amount of axes.

6 Conclusions

Many ways to modify motions were found in the literature survey. Three modification were implemented and their effects on the emotional content of captured motions were evaluated with a questionnaire. Modifying the posture did help to create the emotions *strong* and *sad*. Changing the length of the motion paths did create *weakness*, *tiredness*, *fear* and *excitement*. However, there is still room for improvement in creating those emotions. The modification to speed and acceleration did not produce *anger* and *relaxedness* in a way that would be obvious to the viewers. The results confirm that *happiness* is an emotion that is very hard to see in motions, if the face of the character is not visible to the viewer. However, *excited* motions can be seen as *happy*.

The questionnaire did give a broad view on the effects of the modifications. Still, some details that the participants saw might have been lost, because the questionnaire was considered quite long and sometimes too repetitive. New ways to evaluate motions should be considered, if we do not want to examine only a set of predetermined emotions.

It became apparent that to create more emotions with modifications more advanced methods are needed. A physics model that could be used to calculate forces that are affecting the character could be very useful in getting the motion look more realistic. Allowing the total time that is used in the motions to change could also be useful, if used in combination with other modifications. Since all the modifications can potentially produce unrealistic motions, future work on the automated detection of errors in motions should not be forgotten.

Many of the papers found in the literature survey state that they used angle based representation of the joints. This might explain part of the problems in the papers as angles are hardly an optimal way to process and modify motion signals. It might be worth the effort to reevaluate the methods using quaternions or direction vectors for the joints. A serious evaluation should be based on a lot of people watching the produced motions, because how people see emotions is not universal. This has not been done in any of the papers found in the literature survey. Part of the problem might be that all of the papers do not define well what their methods are intended to do. Many of the papers refer to the results only as changes in the style of the motions.

Modifying motions to produce changes in the emotional content could be very helpful in games. Capturing a lot of motions is not cheap, therefore modifications could lower the price of using motion capture in games. Quality of the performance by amateur actors can vary greatly. With modifications the end result is less depended on the talent of the actors as suitable emotions can be added to the performance later on. Modifications to emotions in motion capture data could allow more freedom in storytelling, because alternative scenes would not necessarily need new motion capture. This feature could be useful in games and animated movies. If the modifications could be made work in real time, they could also be used in communication between people.

In the end modifying motions cannot accomplish everything, therefore other methods must also be used when making a complete animation. Symbolic gestures cannot be created with simple modifications. Collecting them in a database might be a good approach. Facial expressions are a major part in many emotions, therefore they also have to be taken into account.

References

Amaya, K. & Bruderlin, A. & Calvert, T. 1996. Emotion from motion. GI '96: Proceedings of the conference on Graphics interface '96. Toronto, Ontario, Canada. Toronto, Ontario, Canada: Canadian Information Processing Society. P. 222-229.

Bruderlin, A. & Williams, L. 1995. Motion signal processing. SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. New York, USA: Association for Computing Machinery. P. 97-104.

Grassia, S. 2000. Motion Editing: Mathematical Foundations. Course notes from course 26, Motion Editing: Principles, Practice, and Promise. SIGGRAPH 2000, Computer Graphics Course Notes, Annual Conference Series. New Orleans, Louisiana, USA.

Hsu, E. & Pulli, K. & Popovic, J. 2005. Style translation for human motion. SIGGRAPH '05: ACM SIGGRAPH 2005 Papers. Los Angeles, California, USA. New York, USA: Association for Computing Machinery. P. 1082-1089.

Gleicher, M. 1998. Retargetting motion to new characters. SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. New York, USA: Association for Computing Machinery. P. 33-42.

Gleicher, M. & Kovar, L. & Schreiner, J. 2002. Footskate cleanup for motion capture editing. SCA '02: Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation. San Antonio, Texas, USA. New York, USA: Association for Computing Machinery. P. 97-104.

Godøy, R. & Jensenius, A. & Leman M. & Wanderley M. 2010. Musical Gestures: Concepts and Methods in Research. In Godøy, R. & Leman M. ed. Musical Gestures: Sound, Movement, and Meaning. New York, USA: Routledge. Ch. 2. P. 12-35.

Lee, J. & Park, J. & Nam, T. 2007. Emotional Interaction Through Physical Movement. Human Computer Interaction. Vol. 4552. P. 401 – 410. Berlin/Heidelberg, Germany: Springer Verlag.

Menache, A. 2000. Understanding Motion Capture for Computer Animation and Video Games. San Diego, California, USA: Academic Press. 238 p.

Morris, D. 1977. Eleet kertovat, ilmeet puhuvat. Helsinki,Finland: Kustannusyhtiö Otava. 320 p. ISBN 951-1-04538-5. Original title: Manwatching: A Field Guide to Human Behaviour.

Morris, D. 1985. Eleitä, ilmeitä, asentoja. Helsinki,Finland: Kustannusyhtiö Otava. 256 p. ISBN 951-1-08923-4. Original title: A Field Guide to the Human Species.

Neff, M. & Fiume, E. 2008. From Performance Theory to Character Animation Tools. In Klette, R. & Metaxas, D. & Rosenhahn, B. ed. Human Motion: Understanding, Modelling, Capture, and Animation. Dordrecht, The Netherlands: Springer.

Neff, M. & Kim, Y. 2009. Interactive editing of motion style using drives and correlations. SCA '09: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New Orleans, Louisiana, USA. New York, USA: Association for Computing Machinery. P. 103-112.

Neuhaus-Klamer, E. & Dadlani, P. 2009. Mogic: Expressing Affective Messages Between Empty Nester and Their Children. Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. 8 p. (DOI: 10.1109/ACII.2009.5349595)

Shapiro, A. & Cao, Y. & Faloutsos, P. 2006. Style components. GI '06: Proceedings of Graphics Interface 2006. Quebec, Canada. Toronto, Ontario, Canada: Canadian Information Processing Society. P. 33-39.

Shoemake, K. 1985. Animating rotation with quaternion curves. SIGGRAPH '85: In Proceedings of the 12th Annual Conference on Computer Graphics and interactive Techniques. New York, USA: Association for Computing Machinery. P. 245-254.

Unuma, M.& Anjyo, K. & Takeuchi, R. 1995. Fourier principles for emotion-based human figure animation. SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. New York, USA: Association for Computing Machinery. P. 91-96.

Witkin, A. & Popović, Z. 1995. Motion warping. SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. New York, USA: Association for Computing Machinery. P. 105-108.

Xiao, Z. & Yang, X. & Zhang, J. 2005. Motion Data Correction and Extrapolation Using Physical Constraints. IV '05: Proceedings of the Ninth International Conference on Information Visualisation. Washington, DC, USA: IEEE Computer Society. P. 571-576.

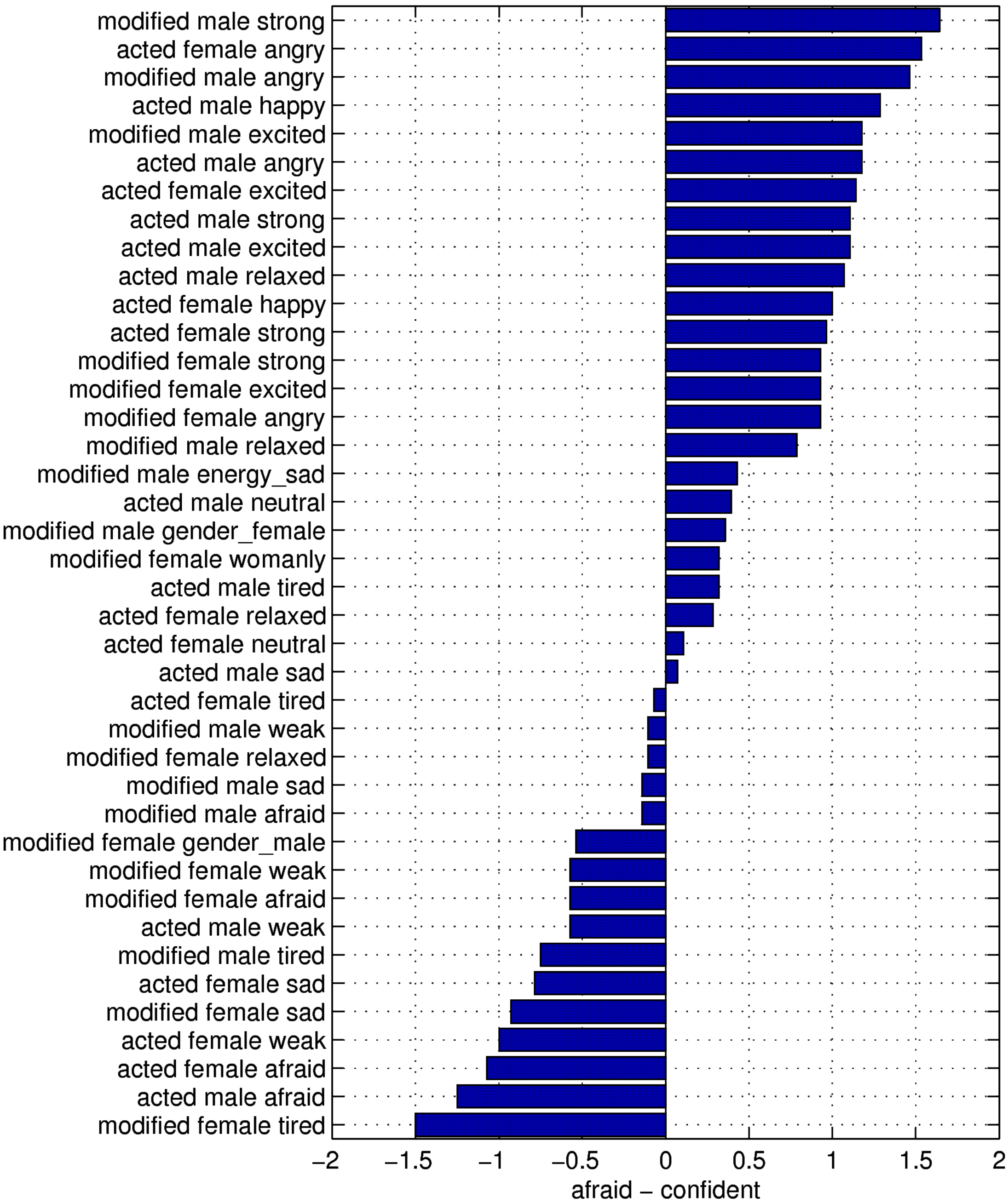
Appendix 1: List of the videos in the questionnaire

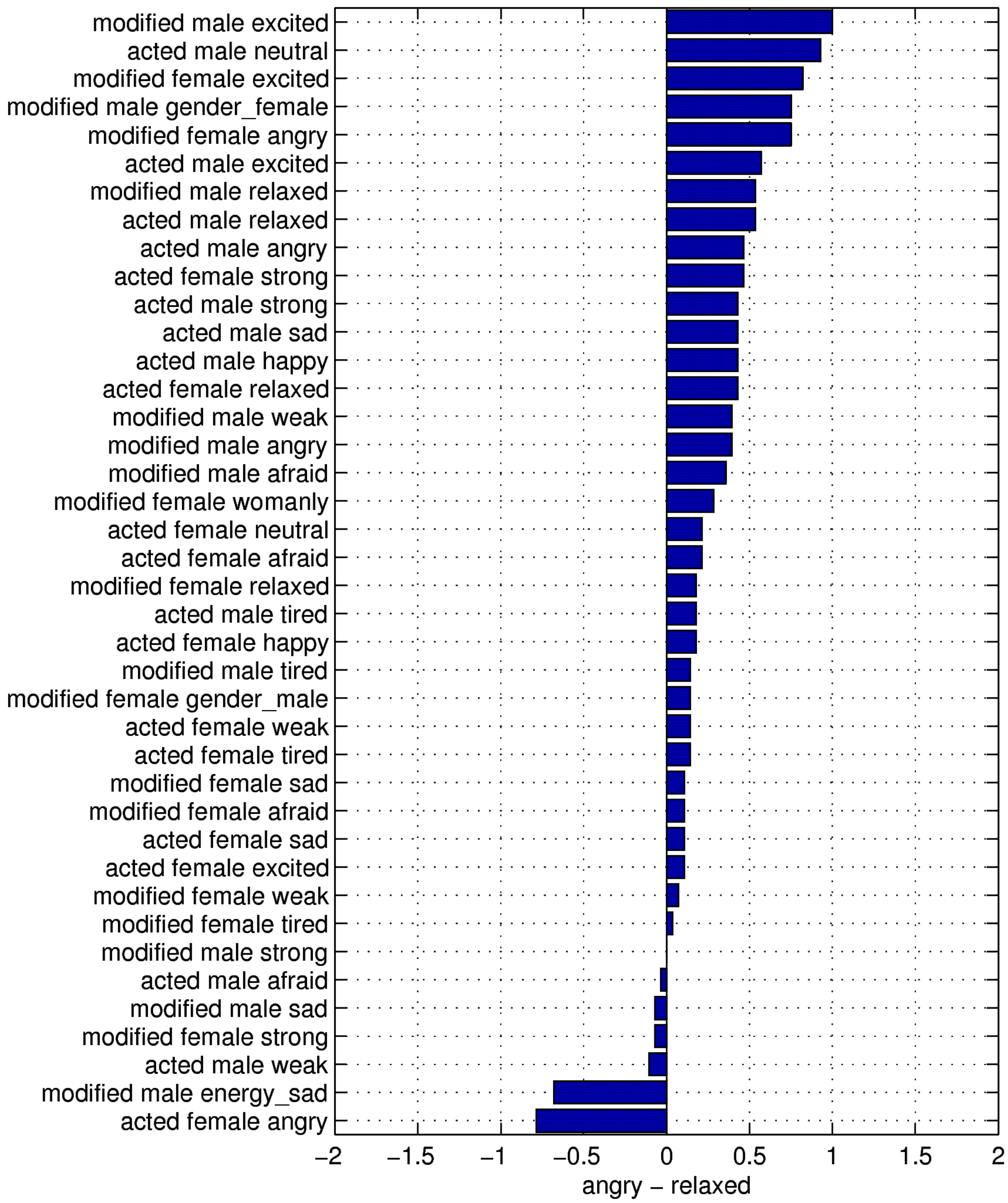
Number of the video	Acted or modified from neutral	Gender of the actor	Emotion intended by the actor or animator
1	acted	female	afraid
2	acted	female	angry
3	acted	female	excited
4	acted	female	happy
5	acted	female	neutral
6	acted	female	relaxed
7	acted	female	sad
8	acted	female	strong
9	acted	female	tired
10	acted	female	weak
11	acted	male	afraid
12	acted	male	angry
13	acted	male	excited
14	acted	male	happy
15	acted	male	neutral
16	acted	male	relaxed
17	acted	male	sad
18	acted	male	strong
19	acted	male	tired
20	acted	male	weak
21	modified	female	afraid
22	modified	female	angry
23	modified	female	excited
24	modified	female	gender_male
25	modified	female	relaxed
26	modified	female	sad
27	modified	female	strong
28	modified	female	tired
29	modified	female	weak
30	modified	female	womanly
31	modified	male	afraid
32	modified	male	angry
33	modified	male	energy_sad
34	modified	male	excited
35	modified	male	gender_female
36	modified	male	relaxed
37	modified	male	sad
38	modified	male	strong
39	modified	male	tired
40	modified	male	weak

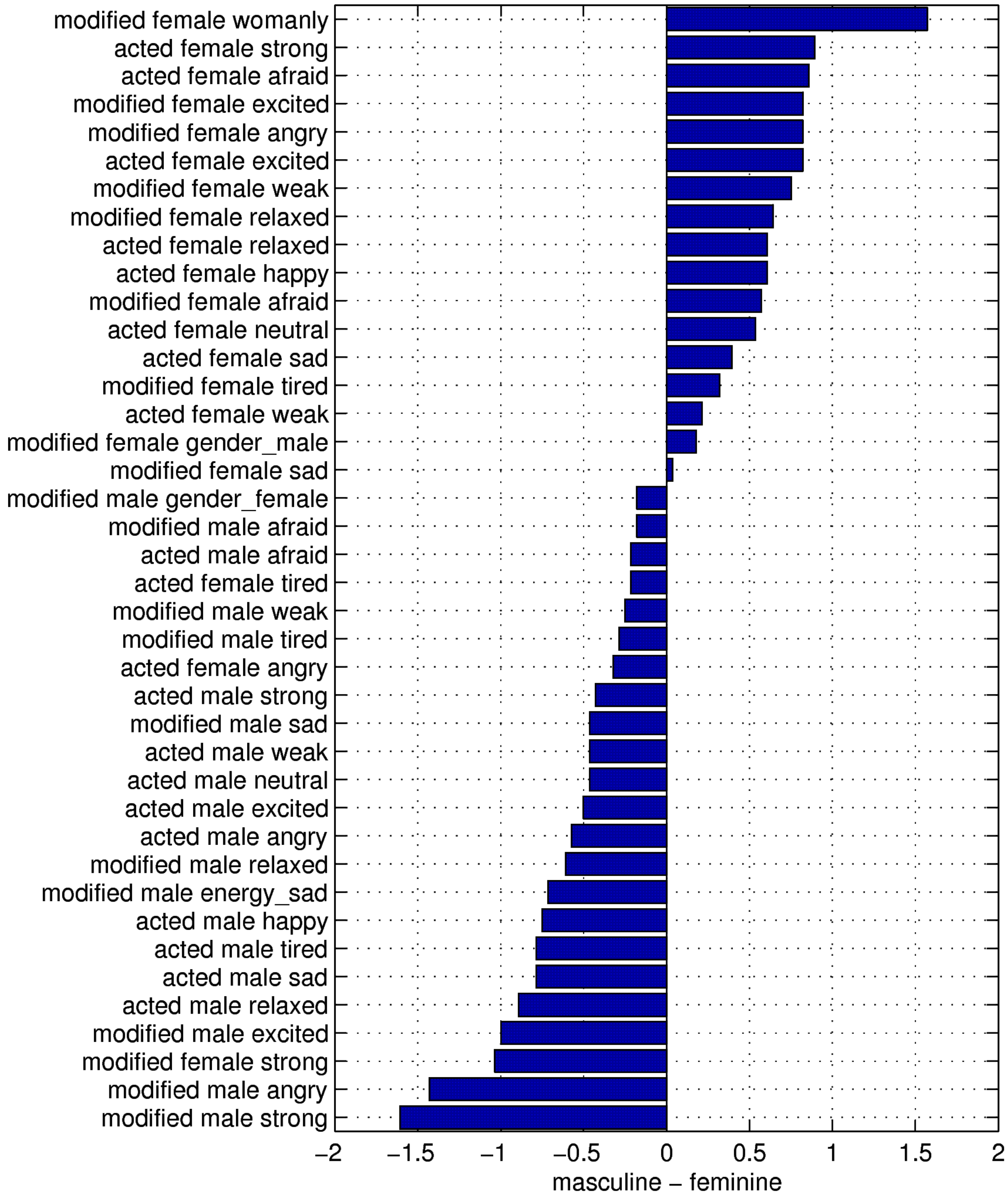
Appendix 2: Free text comments from the participants of the questionnaire

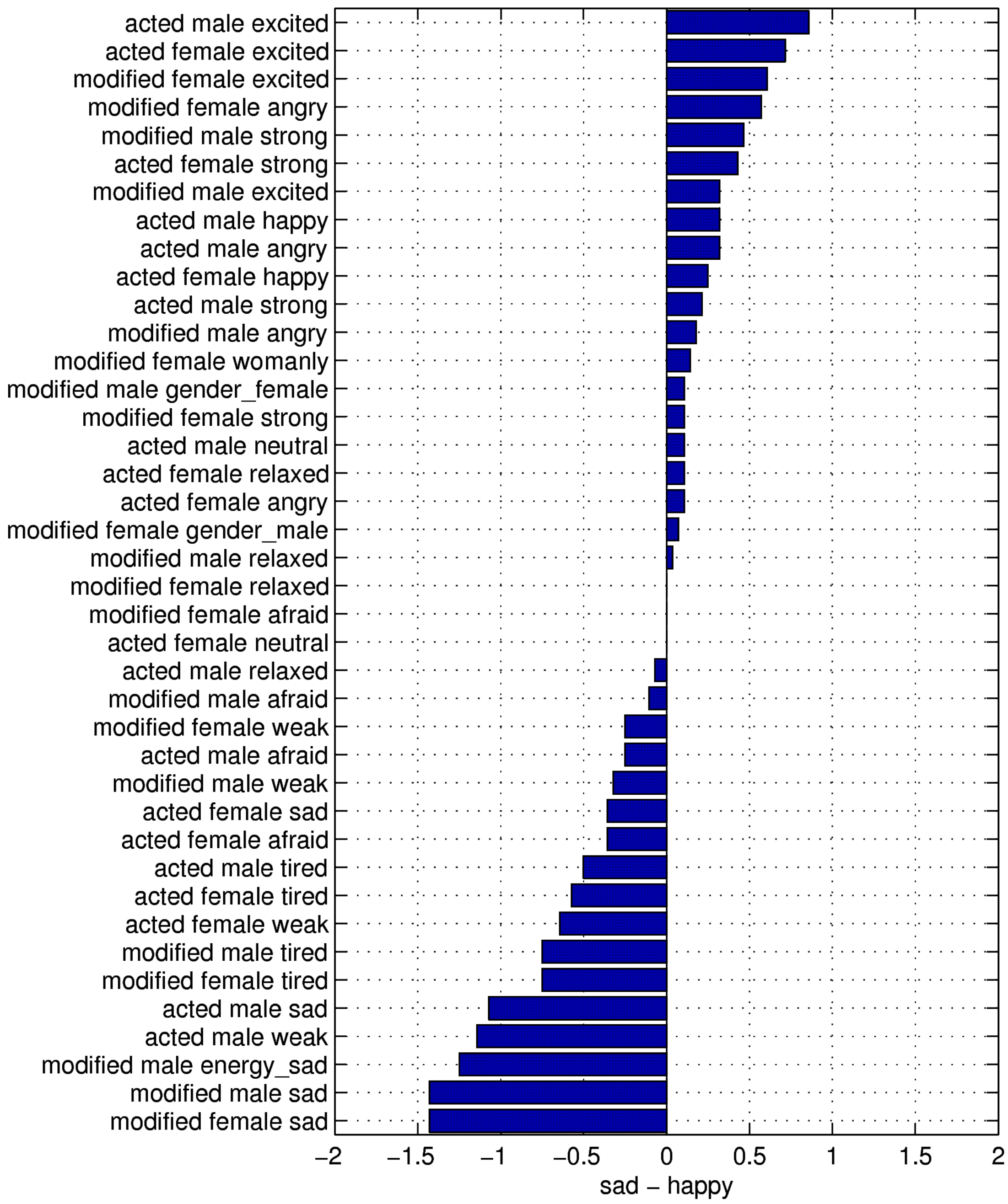
Number of the video	Comment in English (original or rough translation)	Comment in Finnish (if originally in Finnish)
4	A little exaggerated, unnatural motion of the left arm	Vähän liitoiteltu, epäluonnollinen, vasemman käden liike
7	Position of the feet, a little apologizing character, at least careful.	Jalkojen asento, hieman anteeksipyytävä hahmo, varovainen ainakin.
8	wobbly, almost drunken walk	horjuvaa, lähes humalaista kävelyä
9	The hand drops quite lazily down after the knock	Käsi valahtaa melkoisen veltosti alas koputuksen jälkeen.
11	careful	
11	Lazy	
11	More anxious than angry.	
11	Old person?	
15	bored	
16	arrogant	
17	disinterested would fit best	
19	bored	
22	The first steps gave me the impression that the character is either drunk or otherwise in a very cheerful mood. :)	Alun askelluksesta jäi mielikuva, että hahmo on joko kännissä tai muuten vaan kovin letkeällä päällä. :)
22	Seems a bit drunk	
23	Walks like a drag queen, too long steps when compared to the rocking hips	Kävelee kuin drag queen, liian pitkä askel noin keinuvaan lantioon
23	She's drunk again	
24	Looks like a robot. If angry is the opposite of relaxed, then this must be angry.	
25	modest	
26	Walks as comatosed as some (many) residents of the student village. “Excuse me that I am alive.”	Kävelee yhtä koomassa kuin jotkut (monet) Teekkarikylän asukit. "Anteeksi että olen olemassa."
26	Looking for his dog?	
26	The bow seem too exaggerated	Tuntuu jotenkin liioitellulta kumartumiselta.
27	The hand do not feel right when compared to the rest to the character	Kädet eivät jotenkin ole suhteessa muuhun olemukseen
30	Position of the legs (tips of feet together) gives slightly feminine impression	Jalkojen asento (jalkaterät yhdessä) antaa hieman naisellisen vaikutelman
30	VERY feminine	
33	looks exaggerated	liioitellun näköinen
36	The pose of the head towards the sky looks funny.	Hassu tuo nokka taivasta kohden asento päällä.
36	Looks to ask someone something	
37	Adjective 'frail' comes to mind	Adjektiivi 'raihnainen' tulee mieleen.

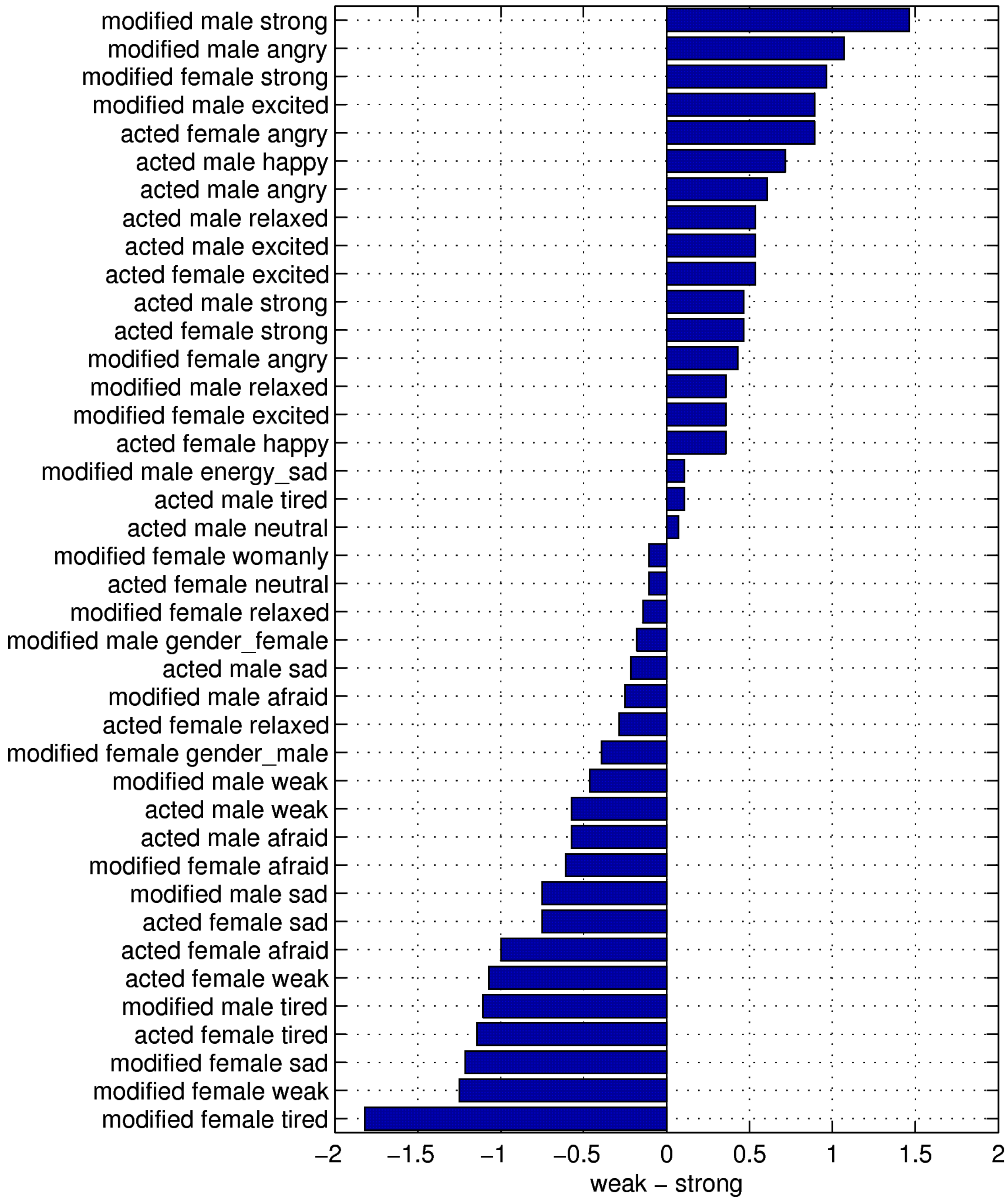
Appendix 3: Videos sorted by the mean values in the axes

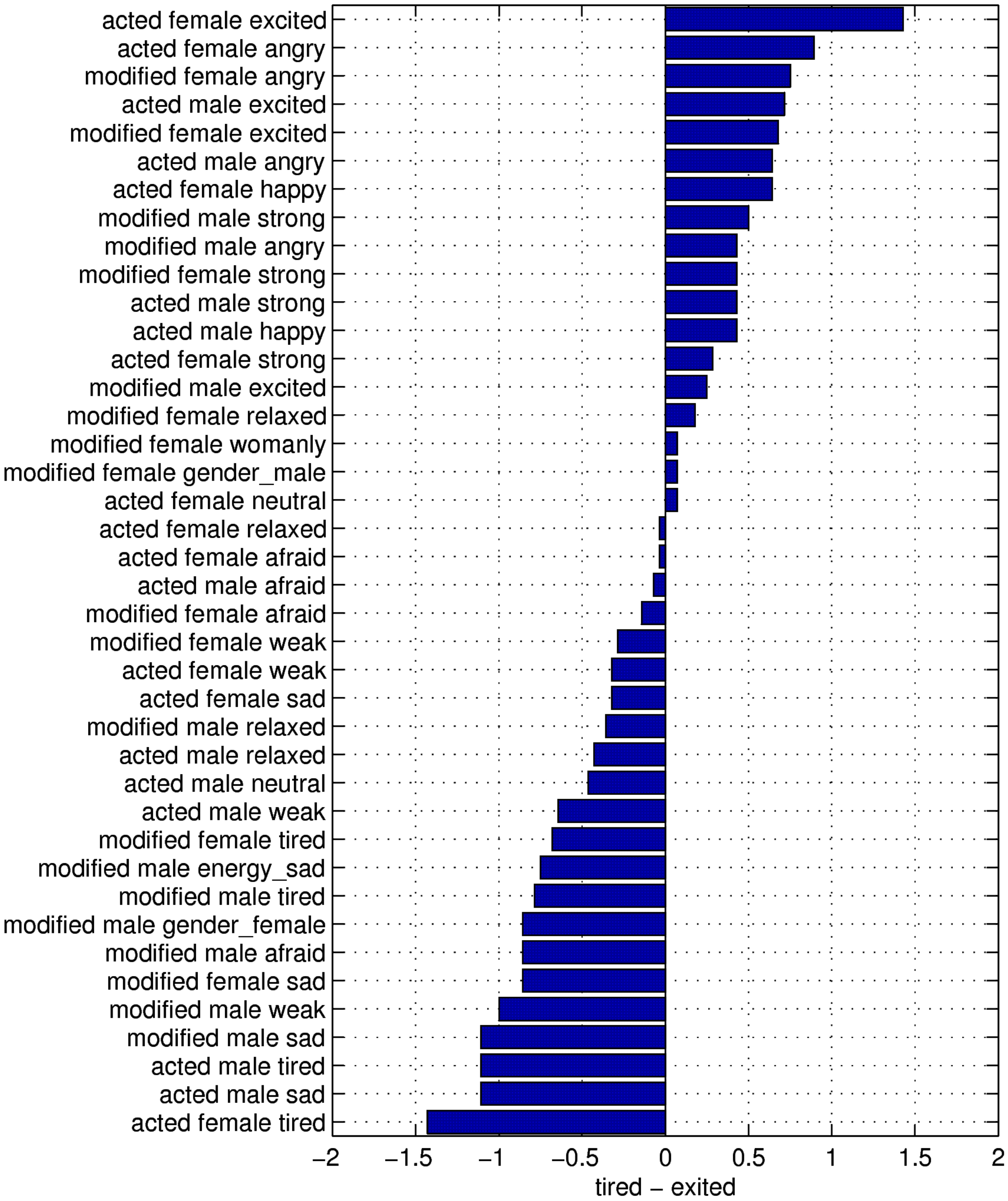



modified male excited
acted male neutral
modified female excited
modified male gender_female
modified female angry
acted male excited
modified male relaxed
acted male relaxed
acted male angry
acted female strong
acted male strong
acted male sad
acted male happy
acted female relaxed
modified male weak
modified male angry
modified male afraid
modified female womanly
acted female neutral
acted female afraid
modified female relaxed
acted male tired
acted female happy
modified male tired
modified female gender_male
acted female weak
acted female tired
modified female sad
modified female afraid
acted female sad
acted female excited
modified female weak
modified female tired
modified male strong
acted male afraid
modified male sad
modified female strong
acted male weak
modified male energy_sad
acted female angry
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
angry − relaxed


modified female womanly
acted female strong
acted female afraid
modified female excited
modified female angry
acted female excited
modified female weak
modified female relaxed
acted female relaxed
acted female happy
modified female afraid
acted female neutral
acted female sad
modified female tired
acted female weak
modified female gender_male
modified female sad
modified male gender_female
modified male afraid
acted male afraid
acted female tired
modified male weak
modified male tired
acted female angry
acted male strong
modified male sad
acted male weak
acted male neutral
acted male excited
acted male angry
modified male relaxed
modified male energy_sad
acted male happy
acted male tired
acted male sad
acted male relaxed
modified male excited
modified female strong
modified male angry
modified male strong
–2
–1.5
–1
–0.5
0
0.5
1
1.5
2
masculine – feminine


acted male excited
acted female excited
modified female excited
modified female angry
modified male strong
acted female strong
modified male excited
acted male happy
acted male angry
acted female happy
acted male strong
modified male angry
modified female womanly
modified male gender_female
modified female strong
acted male neutral
acted female relaxed
acted female angry
modified female gender_male
modified male relaxed
modified female relaxed
modified female afraid
acted female neutral
acted male relaxed
modified male afraid
modified female weak
acted male afraid
modified male weak
acted female sad
acted female afraid
acted male tired
acted female tired
acted female weak
modified male tired
modified female tired
acted male sad
acted male weak
modified male energy_sad
modified male sad
modified female sad
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
sad − happy


modified male strong
modified male angry
modified female strong
modified male excited
acted female angry
acted male happy
acted male angry
acted male relaxed
acted male excited
acted female excited
acted male strong
acted female strong
modified female angry
modified male relaxed
modified female excited
acted female happy
modified male energy_sad
acted male tired
acted male neutral
modified female womanly
acted female neutral
modified female relaxed
modified male gender_female
acted male sad
modified male afraid
acted female relaxed
modified female gender_male
modified male weak
acted male weak
acted male afraid
modified female afraid
modified male sad
acted female sad
acted female afraid
acted female weak
modified male tired
acted female tired
modified female sad
modified female weak
modified female tired
–2
–1.5
–1
–0.5
0
0.5
1
1.5
2
weak – strong


acted female excited
acted female angry
modified female angry
acted male excited
modified female excited
acted male angry
acted female happy
modified male strong
modified male angry
modified female strong
acted male strong
acted male happy
acted female strong
modified male excited
modified female relaxed
modified female womanly
modified female gender_male
acted female neutral
acted female relaxed
acted female afraid
acted male afraid
modified female afraid
modified female weak
acted female weak
acted female sad
modified male relaxed
acted male relaxed
acted male neutral
acted male weak
modified female tired
modified male energy_sad
modified male tired
modified male gender_female
modified male afraid
modified female sad
modified male weak
modified male sad
acted male tired
acted male sad
acted female tired
–2
–1.5
–1
–0.5
0
0.5
1
1.5
2
tired – exited